An investigation into the role of chromatin modifying elements on the production of recombinant antibodies from CHO Cells

Thesis

UNRESTRICTED.

AN INVESTIGATION INTO THE ROLE OF CHROMATIN MODIFYING ELEMENTS ON THE PRODUCTION OF RECOMBINANT ANTIBODIES FROM CHO CELLS

Fay Louise Saunders, BSc (Hons)

A thesis submitted in partial fulfilment of the requirements of the Open University for the degree of Doctor of Philosophy

November 2010

Department of Biology
UCB, 216 Bath Road, Slough,
Berkshire, SL1 3WE, U.K

Department of Medical and Molecular Genetics
King's College London School of Medicine
SE1 9RT, UK

Date of Submission: 4 November 2010
Date of Award: 11 March 2011

Abstract

Stable cell line generation requires the gene of interest to become stably integrated into the host cells genome. The generation of stable cell lines by random integration produces large variation in clonal expression and stability that is due, in part, to the nature of the chromatin it has integrated into.

Elements have been identified that can modify the chromatin structure and have been shown to protect transgene expression from the effects of the neighbouring chromatin environment. The aim of the study was to investigate whether the identified chromatin structure modifying elements (UCOE, MAR, STAR and HS4) could increase the level and stability of antibody expression in stable cell lines in the mammalian expression systems used at UCB.

A series of expression constructs were generated to assess the effect of the chromatin structure modifying elements on antibody expression in stable cell lines. Initially seven different vectors for each element that possessed different combinations of the individual elements were assessed as pooled and clonal stable cell populations which indicated that the elements increased antibody expression by varying amounts and optimal vector configurations were different for the elements.

The optimised vectors were subsequently used in a comparative side-by-side study with expression levels from pooled and stable cell lines being analysed. A subset of clones harbouring each of the five test constructs (Ab535 control, 1.5kb A2UCOE, MAR X_S29, STAR 7 and cHS4 tandem) derived from the pooled stable cell lines

were genetically characterised to determine copy number and clonal diversity. Clones were cultured long-term in the absence of selection pressure which resulted in a decrease in expression being observed in all of the clones. The decline in antibody expression was accompanied by a decrease in mRNA levels which was not caused by a loss of either HC or LC gene copies from the genome. Analysis of DNA methylation patterns across the hCMV-MIE promoter regions demonstrated that this epigenetic regulatory mechanism was involved in the silencing of a number of clones which exhibited a decrease in productivity.

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Paul Stephens. His support and guidance throughout has been invaluable. Thanks also to Dr. Michael Antoniou at King's College London for his scientific advice throughout the project and also for the critical reviewing of this thesis. I am also grateful to Dr. Martyn Robinson, not only for the opportunity to undertake a PhD at UCB but also for the support throughout the project.

Thank you to Prof. Darren Griffin who allowed me to carry out the FISH analysis in his lab at the University of Kent and a very big thank you to Dr. Martin Volker for his time and patience in teaching me this method, it was great for something to work first time! Thanks also to Grace Smith for carrying out the automated miniprepping which was invaluable to completing the bisulphite sequencing.

Thank you to my friends and colleagues at UCB, you made my time in the lab, especially the long hours spent in tissue culture a much more enjoyable experience, and kept me smiling during the more challenging times!

I would also like to thank Mum and Dad; I don't think I would have made it this far in my career without your love and support. Finally, a big thank you to my husband Simon, your complete belief in me has meant so much and I really couldn't have done this without you. I also apologise for having spent the first few months of married life writing this thesis!

Abbreviations

Ab	Antibody
amp	Ampicillin
AP	Alkaline Phosphotase
ATCC	American Type Culture Collection
BHK	Baby Hamster Kidney
BLAST	Basic Search Alignment Tool
bp	Base Pairs
BSA	Bovine Serum Albumin
C	Centigrade
CCD	Charged Couple Device
cDNA	Complimentary Deoxyribonucleic Acid
CD	Chemically Defined
CGI	CpG Island
ChIP	Chromatin immunoprecipitation
CHO	Chinese Hamster Ovary
CSME	Chromatin Structure Modifying Element
CIP	Calf Intestinal Phosphotase
Da	Dalton
DAPI	4', 6-diamidino-2-phenylindole
DGV	Double Gene Vector
DHFR	Dihydrofolate Reductase
DIG	Dioxygenin
DNA	Deoxyribonucleic Acid
DNase	Dioxyribonuclease
dATP	Deoxyadenosine Triphosphate
dCTP	Deoxycytosine Triphosphate
dGTP	Deoxyguanosine Triphosphate
dTTP	Deoxythymosine Triphosphate
dH2O	Distilled Water
EDTA	Ethylenediamine Tetra-Acetic Acid
ELISA	Enzyme –linked Immunosorbent Assay

FAM	6-carboxyfluorescein
FDA	Food and Drink Administration
FISH	Fluorescence In-Situ Hybridisation
Fc	Fragment Crystallisable
FCS	Foetal Calf Serum
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GS	Glutamine synthatase
g	Gram
gDNA	Genomic Deoxyribonucleic Acid
HAT	Histone Acetyltransferase
HC	Heavy Chain
hCMV-MIE	Human Cytomegalovirus with modified immediate early region
HDAC	Histone Deacetylase
HEK	Human Embryonic Kidney
HMT	Histone Methyltransferase
HRP	Horse Radish Peroxidase
cHS4	Chicken Hypersensitive site 4
Ig	Immunoglobulin
IVC	Time Integral of Viable Cell
kb	Kilobase
kDa	Kilodalton
l	Litre
LB	Luria Broth
LC	Light Chain
mAb	Monoclonal Antibody
MAR	Matrix Attachment Region
m	Milli
M	Molar
MSX	Methionine sulphoximine
mRNA	Messenger RNA
n	Nano
neo	Neomycin phosphotransferase
PAC	P1 artifical chromosome
pmoles	Picomoles

PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
PTM	Post Translational Modification
Q-PCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNase	Ribonuclease
rpm	Revolutions Per Minute
RT	Reverse Transcription
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
S.O.C	Super Optimal broth with Catabolite repression
SSC	Saline-Sodium Citrate
STAR	Stabilising Anti-Repressor
SV40	Simian Virus 40
TAE	Tris –acetate EDTA
THF	Tetrahydrofolate
TMB	3,3,5,5'-Tetramethylethylenediamine
TSA	Trichostatin A
UCOE	Ubiquitous Chromatin Opening Element
UV	Ultraviolet
v/v	Volume/Volume
V/cm	Volts per centimetre
VH	Variable Heavy
VL	Variable Light
w/v	Weight per volume
µ	Micro
µF	Micro-faraday

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Recombinant antibody production in mammalian cells

Biopharmaceuticals are one of the fastest growing classes of therapeutics with large numbers in clinical development. An estimated 2,500 biopharmaceutical drugs are in the discovery phase which represents 44% of all drugs in development (Lawrence, 2005). Since the 1980's over 165 recombinant protein therapeutics (such as hormones, growth factors, therapeutic enzymes, vaccines and monoclonal antibody based products) have been approved and this market is expected to reach $70 billion dollars by 2010 (Walsh, 2006).

Since the 1990's antibodies have emerged as an important class of biopharmaceuticals. As drugs, antibodies have many advantages, including having a high target selectivity, high affinity and relatively low toxicity. The United States Food and Drug Administration (FDA) have approved over 20 monoclonal antibodies (mAbs) and several hundred candidates are in development for clinical use in the treatment of various diseases including cancers, immune disorders and infections (Beck *et al.*, 2009). This has led to revenues in 2009 of over $35 billion worldwide (Visiongain, 2010).

Although *E.coli* has been utilised in producing a small number of therapeutic antibodies, the majority are produced in mammalian systems. These systems are technically complex, slow and expensive compared to *E.coli* but are preferred for the production of large, complex proteins as they can manage the complex protein folding and assembly and post-translational modifications (PTMs) required for their correct biological activity (Wurm, 2004).

Initially the production of recombinant monoclonal antibodies (mAb) was limited to the use of hybridomas. Hybridomas are generated by fusing a short-lived specific antibody-producing B cell with an immortalised myeloma cell (Kohler and Milstein, 1975) so that the antibody is continually secreted. However, antibody titres from hybridomas are low (<100mg/L) and they cannot reach a high biomass in bioreactors. This means that they are not suited to large scale manufacturing and instead research focused on stably transfected mammalian cells. A number of different cell lines have been utilised in recombinant antibody production including Baby Hamster Kidney (BHK), Chinese Hamster Ovary (CHO), Human Embryonic Kidney 293 (HEK293), and NS0 cells. However, it is CHO cells that have become the most commonly used host with nearly 70% of recombinant protein therapeutics being made in this line (Jayapal *et al.*, 2007).

1.1.1 Chinese Hamster Ovary (CHO) cells

The original CHO cell line was derived in 1957, when a primary culture of ovarian cells from a Chinese hamster (*cricetulus griseus*) was immortalised (Puck *et al.*, 1958). Further mutagenesis studies resulted in different CHO strains being derived. Initially, the CHO-K1 strain, a glycine dependent sub-clone, was isolated. Further mutagenesis of CHO-K1 cells generated CHO-DXB11 (also referred to as CHO-DUKX or CHO-DUK-XB11), which was deficient in dihyrdrofolate reductase (DHFR) activity (Urlaub and Chasin, 1980). CHO-DXB11 cells have a deletion in one of the *dhfr* alleles and a missense mutation in the other. Further work by Urlaub *et al.* (1983) mutagenised a proline dependent derivative of the original CHO cell line (CHO-pro3$^-$) to generate CHO-DG44 which has deletions in both *dhfr* alleles. The

ability of the cell lines to grow routinely *in-vitro* allowed them to be used in many pioneering experiments and demonstrated their ability to be used as a host for recombinant protein production (Kaufman & Sharp, 1982; Ringold *et al.*, 1981; Scahill *et al.*, 1983). The first recombinant therapeutic protein from a CHO derived cell line to be approved by the FDA was tissue-type plasminogen activator (Activase®) in 1987. However, it was not until 1997 that the first CHO derived recombinant mAb, rituximab, a chimeric anti-CD20 mAb, came to market (Jayapal *et al.*, 2007).

A number of reasons have contributed to CHO cells being the preferred host for recombinant protein production. Perhaps most importantly, CHO cells are readily amenable to genetic manipulation, which allows the introduction of foreign DNA into the cell and subsequent expression of the desired protein. The choice of host cell also has an impact on the product characteristics and yields. CHO cells have been shown to produce IgGs that possess glycoforms which closely resemble human IgG profiles (Jefferis, 2005). Product safety is another important factor and CHO cells have been shown to have a good safety profile, with only a small number of human pathogenic viruses being able to replicate in this cell line (Wiebe *et al.* 1989) and they have a history of regulatory approval for recombinant protein production. Another advantage of CHO cells is that they can be cultured in suspension and can be readily scaled to 20,000L. A large amount of research has focused on increasing yields from industry established mammalian cells so that the cost of production can be decreased. In the mid 1980s, a typical batch culture production run, with no feeding regimes, would last for approximately 7 days with cell densities peaking at approximately 1-2 million cells/ml leading to product titres of approximately 50-100mg/l (Wurm, 2004).

Optimisation of the host cell has since taken place and cells have been selected that grow to high densities of approximately 10-15 million cells/mL. In addition, there have been improvements in media composition and feeding strategies which have resulted in cells remaining viable for longer cultivation periods. These improvements have led to a significant improvement in product titres with cell lines expressing in excess of 5g/L now being reported (Birch and Racher, 2006). In order to achieve these levels of antibody expression, a clonal cell line must be generated which stably expresses the antibody at a high level. Deriving a clone with these characteristics is still an empirical process and large numbers need to be initially selected and screened in order to isolate a clone with the desired characteristics.

1.1.2 Stable cell line generation

The generation of recombinant antibody producing cell lines follows a defined multistep process. This begins with the molecular cloning of the gene of interest (GOI) into a mammalian expression vector. Vector systems must contain all the elements for efficient expression of the antibody genes, including transcriptional regulatory elements (e.g. promoters, enhancers), kozak translation initiation sequences and polyadenylation signals. Strong promoters/enhancer sequences are commonly used to drive expression, which are frequently of viral origin such as the human cytomegalovirus major immediate early (hCMV-MIE) promoter (Boshart *et al.*, 1985) or derived from genes that are highly expressed in mammalian cells such as elongation factor 1 alpha (EF1α) (Running Deer and Allison, 2004). In addition, vectors should also contain at least one intron. The presence of an intron and appropriate splice sites are important for efficient mRNA processing and export from the nucleus to the cytoplasm (Le Hir *et al.*, 2003). Protein production can also be

improved by optimising the coding region sequences for efficient codon usage and to remove motifs that could lead to mRNA instability or undesirable secondary structure (Kalwy *et* al., 2006). In the generation of stable transfected cell lines, cells that contain the vector must be distinguished from those that do not possess the GOI. This is usually achieved by the addition of a selectable marker, which is either included in the expression vector or co-transfected on a separate plasmid.

Selection markers can either be based on dominant or recessive selection. Dominant systems are usually based on the inclusion of a gene which encodes an enzyme that confers antibiotic resistance. As a result cells that have been stably transfected will survive in the presence of antibiotics in the culture medium. An example of this is the neomycin selection system. Mammalian cells are sensitive to the aminoglycoside antibiotic geneticin (G418), as this inhibits protein synthesis. However, incorporation of the bacterial aminoglycoside gene *neomycin phosphotransferase* (*npt*) into the expression vector allows cells which have been transfected to survive in the presence of G418 (Southern and Berg, 1982). Other commonly used antibiotics are Blastocydin, Hygromycin, Puromycin and Zeocin.

Recessive selection systems restore a critical enzyme or protein to the cell that is required for cell survival. The two most commonly used are the dihydrofolate reductase (DHFR) and glutamine synthetase (GS) selection systems. DHFR catalyses the conversion of folic acid, to tetrahydrofolate (THF). $DHFR^-$ cells cannot produce glycine, hypoxanthine and thymidine, and therefore cannot survive in media lacking these components. Introduction of a functional copy of the DHFR gene restores the cell's function to produce these nutrients. Clonal selection can then be performed by

growing the cells in medium lacking glycine, hypoxanthine and thymidine (Lucas *et al.* 1996). The GS selection system works in a similar manner. The GS enzyme catalyses the conversion of glutamate and ammonium to glutamine. When glutamine is absent in the culture medium, the activity of the GS enzyme is essential for cell survival. The ability to grow in glutamine free medium can be conferred by the transfection of a functional GS gene (Bebbington *et al.*, 1992). This system works particularly well in NS0 cells, which lack endogenous GS activity and cannot survive without glutamine in the culture medium. CHO cells in contrast have endogenous GS activity and can survive in medium lacking glutamine. For the selection system to be successful in this cell line a GS inhibitor, methionine sulphoximine (MSX) must be added so that only transfectants with additional GS activity can survive.

The cells that survive selection pressure are highly heterogenous in terms of cell growth and specific productivity. This means that a large number of individual clones need to be isolated and evaluated in order to obtain cell lines with the desired characteristics, which can be a lengthy and time consuming process.

Large numbers of clones need to be screened because integration of DNA into the genome is a random and inefficient process, with a frequency of only 1 event in 10,000 (Gorman and Bullock, 2000) and DNA can become integrated at any position within the genome. Integration does not always guarantee prolonged transgene expression with studies showing that stably transfected cell lines can exhibit a marked progressive reduction in transgene expression after a few tens of generations (Barnes *et al.*, 2001; Kim *et al.*, 1998; Forrest-Owen *et al.*, 2005). This will have an impact on the process development of a therapeutic cell line. For a successful manufacturing

cell line, the expression of the recombinant gene must be constant over a long period in the absence of selection pressure. Cell line stability is generally required over a period of at least 60 generations beyond the production of a manufacturing working cell bank which ensures sufficient time for inoculum expansion, the production process and a 20 generation margin for safety (Brown *et al.*, 1992).

1.2 Position effects and chromatin

The chromatin structure surrounding the site of integration can influence transgene expression in a positive or negative way as the chromatin is organised in two distinct structures, called euchromatin and heterochromatin (whose properties will be discussed further in later sections). Therefore, the site of integration will effect how well a transgene is expressed; integration of a transgene into the highly condensed, silenced heterochromatin will result in little or no expression, whereas integration into the more open, active euchromatin will generally result in more prolonged expression of the transgene. A related process, known as Position Effect Variegation (PEV) is a type of position effect which is due to heterochromatin being able to silence euchromatic genes when they are brought into close proximity by a chromosomal rearrangement or transposition. The condensed structure of heterochromatin can spread into euchromatic regions, causing them to assume the morphology of heterochromatin and become transcriptionally inactive. The silencing of genes which normally express a particular protein results in a variegated phenotype being observed (Wallrath, 1998). PEV was first observed in *Drosophila* where the *white*$^{+}$ gene (required for red eye pigmentation) was placed near centric heterochromatin by a

chromosomal inversion. As a result a variegated phenotype was observed which resulted in a red and white mosaic eye pattern (Muller and Altenberg, 1930).

1.3 Chromatin structure and composition

The mammalian genome is composed of approximately 3 x 10^9 bp DNA that is almost 2m in length, which needs to be compacted into the small (5-10nm diameter) volume of the nucleus. This is achieved by arranging DNA into chromosomes. Chromosomes are comprised of chromatin, where the principal components are DNA and histones. Histones are small basic proteins which consist of a central globular domain and N- and C- terminal tails. The fundamental building block of chromatin is the nucleosome (Van Holde *et al.*, 1974). The nucleosome is made up of 147bp of DNA wrapped around an octamer of four histone dimers (H2A, H2B, H3 and H4) (Figure 1.1). The central globular core consists of a histone fold domain, which is a structural motif which allows interaction with other histone proteins (Arents *et al.*, 1991). Each histone octamer is composed of two H3-H4 histone dimers that are bridged together as a stable tetramer. This is then flanked by two separate H2A-H2B dimers (Luger *et al.*, 1997). DNA is wrapped nearly twice around the octamer core and this provides the first level of compaction, resulting in a 40-fold decrease in the linear length (Geyer and Parnell, 2000). The resultant nucleosomes form 11nm fibres that resemble beads on a string (Olins and Olins, 1974). Further compaction is provided by the addition of a linker histone, H1, forming the chromatosome. H1 binds to the nucleosome at the DNA entry and exit points and acts to protect the linker DNA (Campos and Reinberg, 2009).

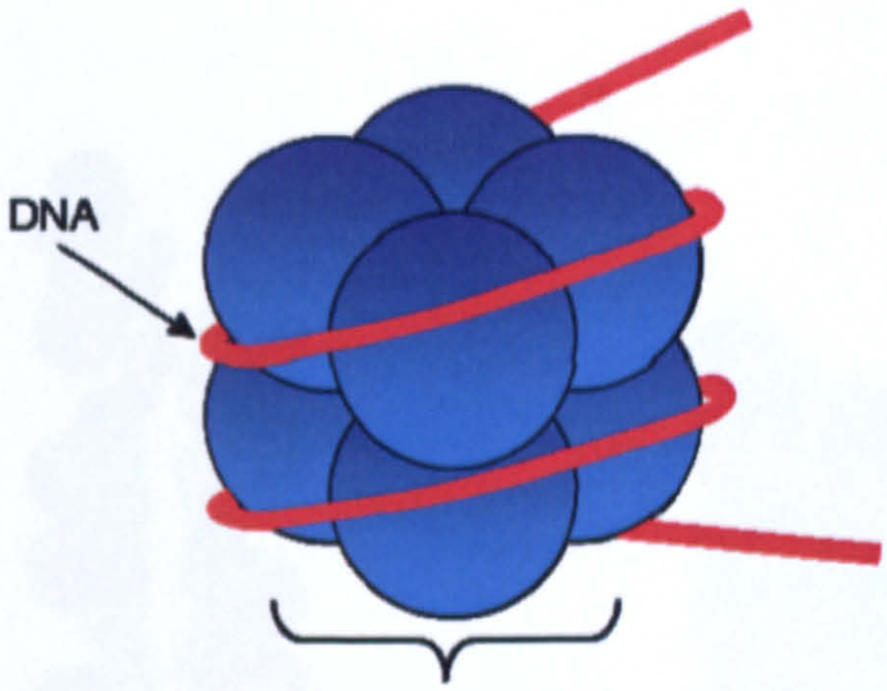


Figure 1.1 The nuclesome
Diagram showing the structure of the nuclesome. DNA is wrapped around a central histone octamer to form a single nuclesome

Further compaction results in what is known as a 30nm fibre, which consists of approximately 6 nucleosomes every 11nm of chromatin fibre (Bassett *et al.,* 2009). This 30nm fibre can then be condensed to varying degrees. Depending on the degree of packing density the chromatin can be split into two distinct classes: Euchromatin and heterochromatin. Figure 1.2 shows schematically the difference between the two classes. Euchromatin is a less condensed 'open' structure in which no further compaction of the 30nm fibre takes place. Heterochromatin is where the 30nm fibre has been compacted further. These regions have a higher density of closely packed nucleosomes.

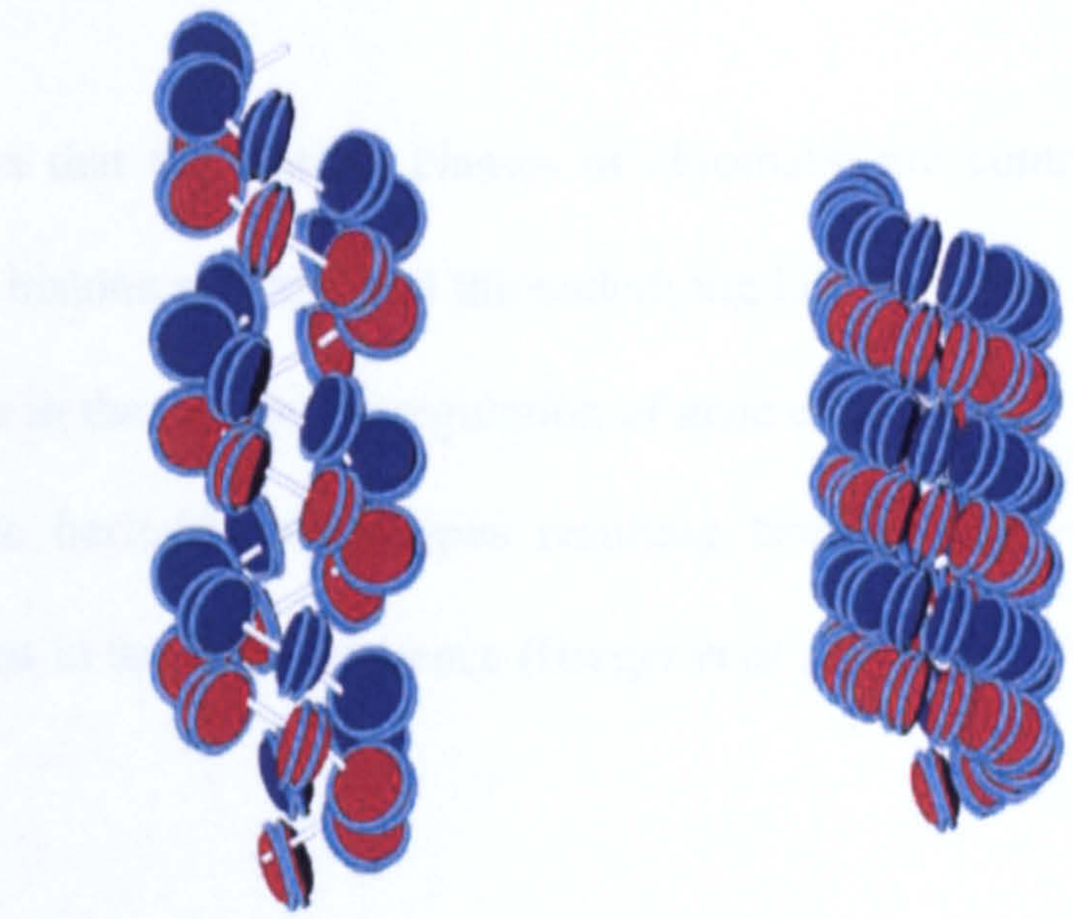


Figure 1.2 Nucleosome packing in euchromatin and heterochromatin
A. In euchromatin nucleosomes are loosely packed in the 30nm fibre. B. Compaction of the 30nm fibre resulting in tightly packed nucleosomes in heterochromatin. (Adapted from Bassett *et al.*, 2009).

The modulation of the structure of the chromatin fibre is critical for the correct regulation of gene expression. The level of compaction determines how accessible the DNA is to regulatory proteins, such as transcription factors and RNA polymerases. In euchromatin, where the nucleosomes are loosely packed the DNA is accessible to these factors. However, in comparison heterochromatin is condensed and closed (Quina et al., 2006). Heterochomatin can be further divided into constitutive and facultative heterochromatin. Constitutive heterochromatin is generally gene poor and forms on repetitive sequences such as satellite centromeric and pericentric repeats. Regions of chromatin that are subjected to developmentally regulated transcriptional silencing are called faculative heterochromatin. An example of facultative heterochromatin is the inactive X-chromosome in female mammalian cells (Kosak and Groudine, 2004).

1.4 Regulation of chromatin structure

Figure 1.3 shows that the distinct classes of chromatin are controlled by chemical modifications to histone proteins and the underlying DNA. These modifications play an important role in the epigenetic regulation of gene expression. Epigenetics can be defined as stable heritable phenotypes resulting from changes in a chromosome without alterations in the DNA sequence (Berger *et al.*, 2009).

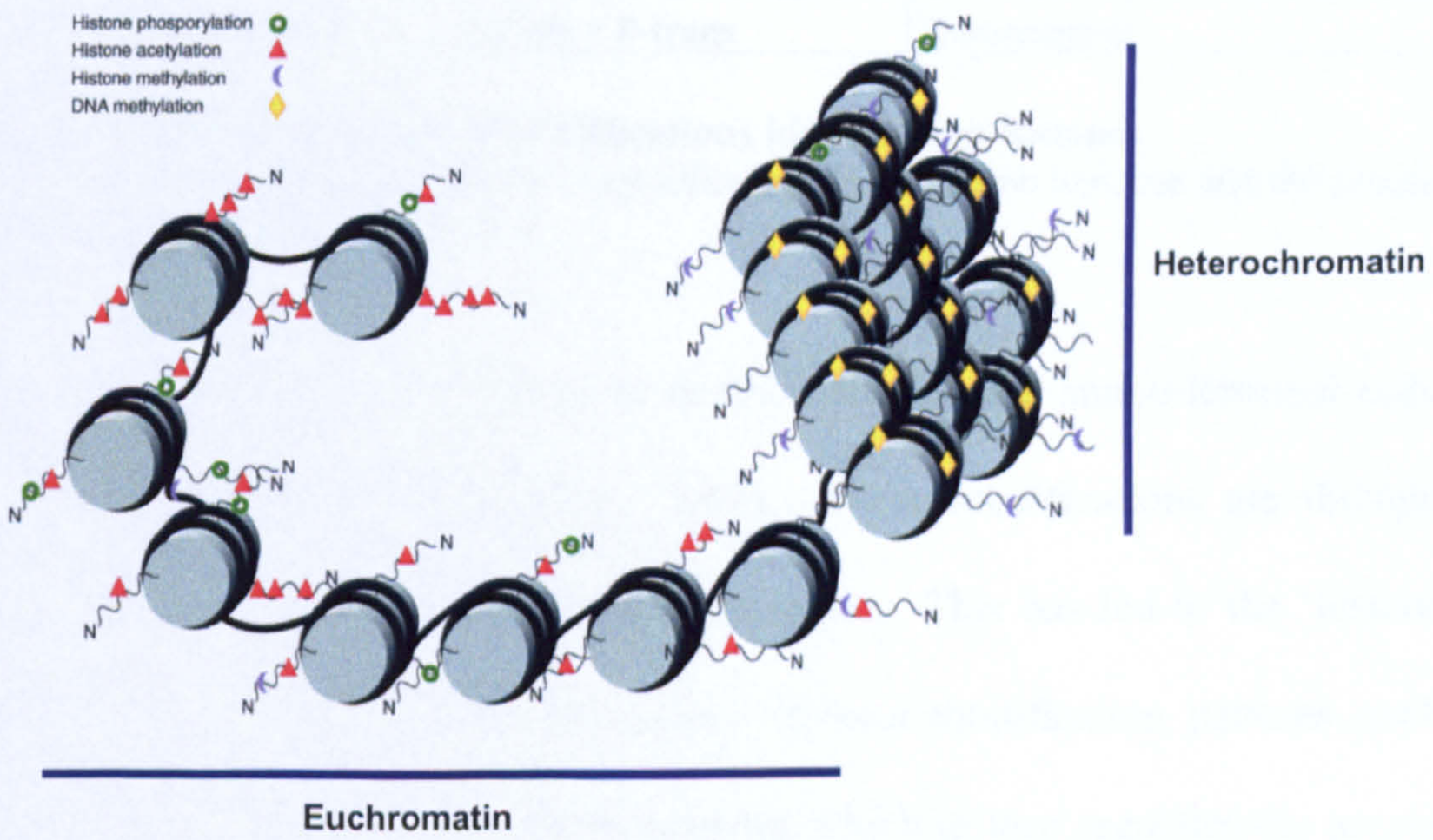


Figure 1.3 Modifications to histones and DNA
The figure shows two hypothetical epigenetic states of open euchromatin (left), which is associated primarily with histone acetylation and phosphorylation patterns, and compact heterochromatin (right), primarily associated with histone methylation and DNA methylation tags. The combination of histone and/or DNA modifications can create a variety of different epigenetic patterns. There are endless options for the combination of epigenetic modifications to provide locus-specific patterns. (adapted from Meyer, 2001)

1.4.1 Covalent modifications of histone proteins

The unstructured amino terminal tails of histone proteins protruding from the core region are subject to post translational modifications (PTMs) (Figure 1.3). A number

of different types of post-translational modifications have been identified and these are summarised in Table 1.

Chromatin Modifications	Residues modified	Functions regulated
Acetylation	K-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	K-me1 K-me2 K-me3	Transcription, Repair
Methylation (arginines)	R-me1 R-me2 R-me3	Transcription
Phosphorylation	S-ph T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP Ribosylation	E-ar	Transcription
Deimination	R > Cit	Transcription
Proline Isomerization	P-cis > P-trans	Transcription

Table 1.1 Different classes of modifications identified on histones
Overview of the different classes of modification identified on histones and the functions they regulate (Kouzarides, 2007)

Over 100 different PTMs occur at distinct sites on the amino-terminal tails of the histone proteins (Bernstein *et al.*, 2007). These modifications are thought to be related to, or even predict transcriptional states. This has led to the 'histone code' hypothesis, which suggests that distinct histone modification patterns control the interaction of chromatin-associated proteins which in turn regulates the access to the associated DNA (Jenuwein and Allis, 2001). The distinct pattern of modifications to the N-terminal tails can act as a marker to recruit non-histone proteins. Depending on the composition of the modifications they can either serve to bind proteins or exclude them from the chromatin. The proteins which bind will themselves further modify the structure of the chromatin. The PTMs can also directly affect chromatin structure; the addition of chemical moieties to the histone tails can disrupt the contacts between nucleosomes and between DNA and the histone octamer core complex which can act to 'unravel' the chromatin (Quina *et al.*, 2006).

One of the most prevalent PTM is the acetylation of lysine residues, occurring on a number of residues in the N-terminal tails of all four core histones. This PTM was first identified over forty years ago (Allfrey *et al.*, 1964) and has long been associated with transcriptional activation. This PTM has a high potential to unfold chromatin, the addition of an acetyl group onto a lysine residue neutralises its positive charge and thereby weakens the affinity of the histone for the negatively charged DNA phosphate backbone. This has been illustrated in a study by Shrogen-Knaak *et al.* (2006). Homogenous acetylation of Lys-16 on Histone H4, a site for acetylation (H4K16Ac), by chemical modification was shown to inhibit the formation of 30nm chromatin fibres when the histone variant was incorporated into nucleosomal arrays. Acetylation can also alter histone:histone interactions between neighbouring nucleosomes (Wolffe and Hayes, 1999) and interactions between histones and regulatory proteins (Edmondson *et al.*, 1996). These modifications alter the chromatin structure which makes it more permissive for transcription.

The transfer of acetyl groups from acetyl coenzyme A to the ε-amino group of lysine residues is catalysed by histone acetyltransferases (HATs). A number of HATs have been identified and many of these are part of larger multisubunit complexes which are recruited to promoters by transcription activators (Utley *et al.*, 1998). HATs can be divided into distinct families due to sequence similarities between family members (Kuo and Allis, 1998). The three main families are Gcn5-related N-acetyltransferase (GNAT), MYST (named after its founding members MOZ-YBF2/SAS3-SAS2-TIP60) and p300/cAMP response element binding protein (CBP) and each appear to have different preferred substrates and distinct functional roles. The HATs that make up the GNAT superfamily are grouped together due to the sequence homology that

they share in four conserved regions spanning over 100 residues (Neuwald and Landsman, 1997) and these proteins interact with transcriptional activators and preferentially acetylate H3K14. The preferred substrate for MYST HATs appears to be histone H4. Although MYST family members have high sequence homology, the functions of the individual HATs are distinct and each one has a distinct role in a wide range of regulatory functions in various organisms (Sterner and Berger, 2000). The p300/CBP proteins were originally identified as transcriptional activators, however they were subsequenctly found to possess HAT activity (Bannister and Kouzarides, 1996).

Conversely, the removal of acetyl groups by histone deactylases (HDACs) results in the hypoacetylation of lysine residues and is associated with transcriptional repression. The removal of acetyl groups restores the positive charge to the histone tail allowing association with neighbouring nucleosomes to form a compact, inaccessible form of chromatin. There are ten structurally related HDACs identified to date that fall into two classes, class I and class II HDACs (Thiagalingam *et al.*, 2003). A third unrelated class of HDACs (class III) consist of homologs of the yeast nicotinamide (NAD) dependent enzymes of the Silent Information Regulator (SIR) family (Landry *et al.*, 2000).

In addition to histone acetylation other histone PTMs are also involved in both transcriptional activation and repression. Each type of modification can induce both positive and negative effects on transcription depending on their location and context.

Methylation can take place on both arginine and lysine residues and is carried out by histone methyltransferases (HMTs). Lysine methyltransferases possess a high degree of specificity compared to HATs. Each HMT modifies a single lysine residue on a single histone (Bannister and Kouzarides, 2005). Methylation of certain residues, including H3K4, H3K36 and H3K79 induce transcription activation whereas methylation of others, such as H3K9, H3K27 and H4K20 induces transcriptional repression (Barski *et al.*, 2007). For example tri-methylation of H3K27 has been shown to induce gene silencing in early development by recruiting repressive Polycomb group (PcG) complexes leading to *de novo* DNA methylation (Ru Cao *et al.*, 2002; Schlesinger *et al.*, 2007). Arginine methylation is also involved in transcriptional regulation, and depending on the location of this PTM it can lead to transcriptional activation (Bauer *et* al., 2002) or repression (Wysocka *et al.*, 2006). Arginine methyltransferases, such as cofactor associated arginine methyltransferase (CARM1) and protein arginine methyltransferase I (PRMT1) are recruited to promoters by transcription factors and can either mono or dimethylate arginine residues on their guanidine nitrogen.

Histone phosphorylation, at Ser-10 on histone H3 plays an important role in transcriptional activation. It is thought that the addition of a negatively charged phosphate group to the N-terminal tail may disrupt electrostatic interactions between the nucleosome and the negatively charged DNA making the underlying genome more accessible to nuclear factors (Cheung *et al.*, 2000). Although this type of PTM is implicated in transcriptional activation it can also have an opposing effect on the chromatin architecture. During mitosis mitotic H3 kinases are activated which

phosphorylates serine residues, including the Ser-10 residue of histone H3 which induces the condensation of the chromosomes (Hendzel *et al*, 1997).

1.4.2 ATP-dependent chromatin remodelling

A second class of chromatin modifying enzymes is the ATP-dependent chromatin remodelers. These are involved in the regulation of chromatin structure and are required for establishing or maintaining a particular transcriptional state. The chromatin remodeling activity can be modulated by other proteins and the enzymes are typically found as subunits of multiprotein complexes. These enzymes contain a catalytic subunit which is related to the Sucrose Non Fermenting (SNF) 2-like family of ATPases (Eisen *et al.*, 1995) and use the energy of ATP hydrolysis to change the location and conformation of the nucleosome without covalent modifications. The SNF2 family of ATP-dependent chromatin remodelers can be further divided into a number of different subfamilies depending on the composition of the surrounding protein subunits, and three major subfamilies have been identified: i) the mating type SWItch/Sucrose Non Fermenting subfamily (SWI/SNF) which contains a bromodomain in addition to the ATPase region, ii) the Imitation of SWI (ISWI) subfamily, which contain a SANT domain (a highly conserved motif whose name comes from the proteins originally identified as containing this motif: Swi3-Ada2-N-cor-TFIIB) and possess fewer subunits compared to the SWI/SNF group and iii) the Mi2/Chromo-Helicase DNA binding (CHD) group, which contain a chromodomain and a DNA binding motif (Vignali *et al*, 2000). HDAC activity has also been found in Mi2/CHD subfamily complexes (Tong *et al.*, 1998). ATP-dependent remodelers are involved in a number of cellular processes including transcriptional regulation (Mizuguchi *et al.*, 1997), replication (Zhou *et al.*, 2005), DNA repair (Van Attikum *et*

al., 2004) and recombination (Fritsch *et al.*, 2004). Some of the complexes rearrange the chromatin so that the DNA is more accessible for transcription, whereas other complexes help to generate a chromatin environment which leads to long term silencing of genes.

1.4.3 DNA methylation

DNA methylation plays an important role in the epigenetic regulation of gene expression. It is found in the genomes of a large variety of organisms, including both prokaryotes and eukaryotes and is associated with transcriptional repression (Bird and Wolffe, 1999). Studies have shown that DNA methylation is involved in a number of biological and developmental processes, such as embryonic development (Li *et al.*, 1992), genomic imprinting (Reik and Walter, 2001), X-chromosome inactivation (Beard *et al.*, 1995) and the suppression of transposon elements (Yoder *et al.*, 1997).

In mammals this DNA modification occurs on cytosine residues and is almost exclusively where the cytosine is immediately 5' to a guanosine. DNA is methylated by the transfer of a methyl group from S-adenosyl-methionine (SAM) to the carbon 5 position of the pyrimidine ring of cytosine. Methylation is carried out by DNA methyltransferases (DNMTs). Structurally the DNMTs can be divided into two domains. The N-terminal region acts as a regulatory domain, whereas the catalytic activity resides in the C-terminal domain. The C-terminal catalytic domains are closely related and these regions are highly conserved amongst all DNMTs from bacteria to mammals (Bestor *et al.*, 1988). There are three different enzymatically active DNMTs identified in mammals which belong to two different families (Bestor, 2000), based on differences in their N-terminal domains (Figure 1.4).

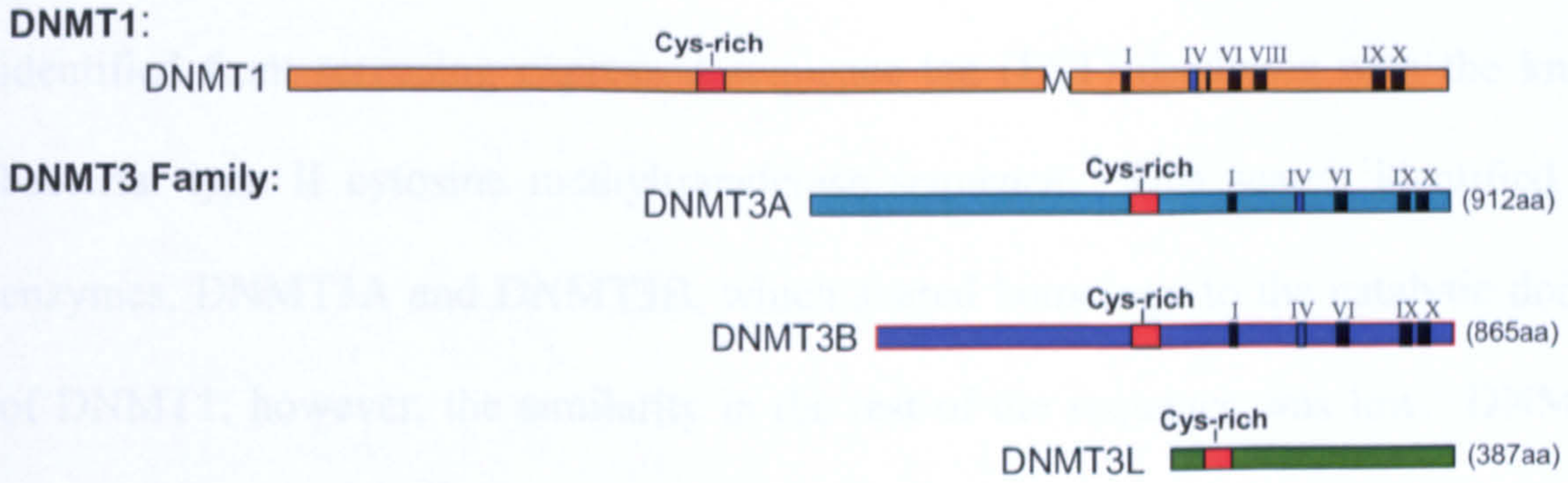


Figure 1.4 The two mammalian DNMT families
The C-terminal catalytic domains of the three enzymes are conserved between the two families; however the N-terminal regulatory domains share no homology. DNMT1 is known as a maintenance DNMT, where as DNMT3a and DNMT3b are *de novo* DNMTs (adapted from Bestor, 2000).

The pattern of DNA methylation is maintained by the maintenance methyltransferase DNMT1. This is the most abundant DNMT in mammals, and is the best characterised, since it was the first eukaryotic DNMT to be cloned (Bestor *et al.*, 1988). DNMT1 has been shown to have a 7-21 fold preference for hemi-methylated DNA and therefore its primary function is believed to be replicating the methylation patterns from the parental to daughter DNA strands following DNA replication (Pradhan *et al.*, 1999). During DNA replication, DNMT1 is brought to the replication fork via the proliferating cell nuclear antigen (PCNA) protein (Chuang *et al.*, 1997) where it can then methylate DNA on the newly synthesized DNA strands. In contrast to the *de novo* DNMTs, DNMT1 contains a large N-terminal region which contains a number of functional domains. These are thought to be responsible for the import of the protein into the nuclei, the co-ordination of replication and methylation during S-phase, and interaction with histone deacetylases. The domain also contains an allosteric DNA binding site which influences the enzymatic activity of the C-terminal domain (Yokochi and Robertson, 2002).

The DNMT3 family is distantly related to DNMT1. This family of DNMTs was identified from screening expressed sequence tag (EST) databases with the known bacterial type II cytosine methyltransferase sequence. This search identified two enzymes, DNMT3A and DNMT3B, which shared homology to the catalytic domain of DNMT1; however, the similarity in the rest of the sequence was low. DNMT3a and DNMT3b are smaller in size than DNMT1 (Figure 1.4) as the N-terminal domains are smaller and have various isoforms (Okano *et al.*, 1998). The DNMT3 family has not been as well characterized biochemically as DNMT1 as they possess a lower catalytic activity. They are highly expressed in embryonic stem (ES) cells, early embryos and developing germ cells. DNMT3a has been shown to have a three fold preference for unmethylated DNA, and therefore it is believed to be a *de novo* DNMT, which is involved in the methylation of cytosine residues during development (Yokochi and Robertson, 2002).

A third member of the DNMT3 family has also been identified based on its sequence similarity to DNMT3a and DNMT3b, called DNMT3L (Aapola *et al.*, 2000). This is expressed specifically in germ cells and is involved in the establishment of patterns of DNA methylation during gametogenesis. On its own it lacks any DNA methyltransferase activity as the protein does not possess the conserved residues involved with catalytic activity. However, it has been shown that interaction with DNMT3a and DNMT3b through the carboxyl-terminal half of the protein results in a marked increase in activity of the two DNMTs. This is irrespective of the DNA sequence and therefore it is believed that DNMT3L acts as a general stimulatory factor for DNMT3a and DNMT3b (Suetake *et al.* 2004). It has also been shown to interact with the amino terminus of histone H3, at the Lys-4 residue (H3K4), however, this only occurs when the lysine residue is unmethylated. It is therefore

believed that DNMT3L recognises unmethylated H3K4 residues and induces *de novo* DNA methylation (Ooi *et al.*, 2007).

1.4.4 DNA methylation and gene repression

DNA methylation can result in the repression of gene expression by two general mechanisms. Firstly, the methylation of cytosine bases can directly inhibit the binding of transcription factors to their DNA binding sites (Watt and Maloy, 1988). Binding of the transcription factors E2F, cAMP Response Element Binding (CREB) and c-myc have all been shown to be disrupted by cytosine methylation (Campanero *et al.*, 2000; Prendergast and Ziff, 1991; Iguchi-Ariga and Schaffner, 1989). Secondly, proteins that recognise methyl-CpG (mCpG) can bind to the methylated DNA and mediate repression. mCpG-binding proteins (MBPs) use transcriptional co-repressor molecules to silence transcription and silence the surrounding chromatin.

The first MBP to be isolated was Methyl CpG binding protein 2 (MeCP2) (Meehan *et al.*, 1989; Lewis *et al.*, 1992). It is comprised of a multidomain polypeptide which contains a mCpG binding domain (MBD) and a transcriptional repression domain (TRD) (Figure 1.5). EST database searches for sequences encoding a conserved MBD domain identified four additional proteins, known as MBD1, MBD2, MBD3 and MBD4 (Hendrich and Bird 1998). A novel MBP protein named Kaiso has also been identified (Prokhortchouck *et al.*, 2001). In contrast to the MBD containing proteins, Kaiso recognizes mCpG using zinc finger domains present at the C-terminus. It mediates repression by associating with nuclear receptor corepressor (N-CoR) complexes which contains HDACs.

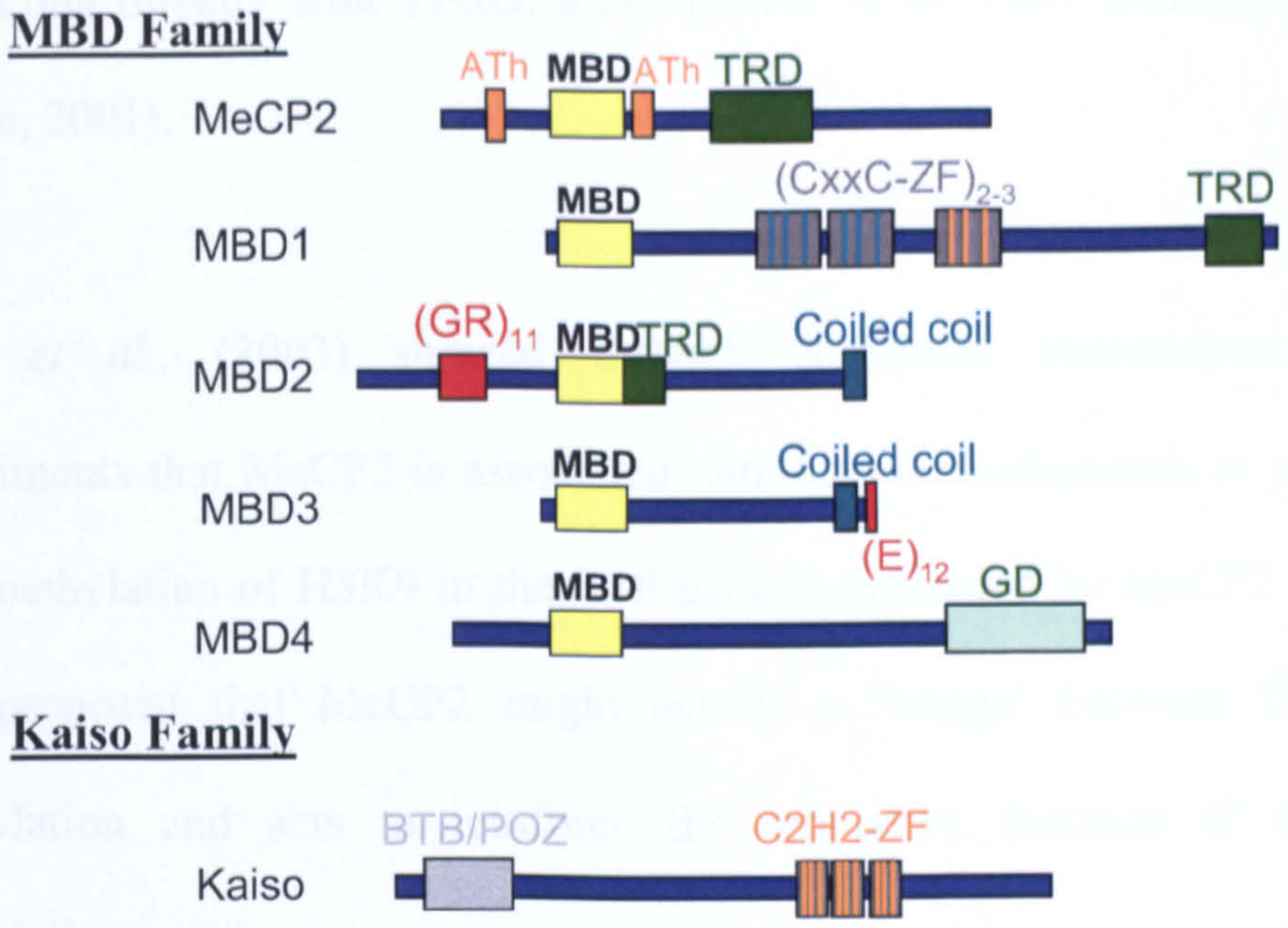


Figure 1.5 Methyl binding proteins
MBP family proteins share a conserved methyl binding domain (MBD), which is required for binding to methylated DNA. MBD3 carries a mutation in the MBD and does not bind to methylated CpG. MeCP2 has two AT-hook motifs (ATh) which can potentially bind AT-rich DNA. MBD1 is characterized by two (or three in some isoforms) CxxC- type zinc fingers. TRD indicates transcriptional repression domain. GD indicates the glycosylase domain of MBD4, which is involved in excision of CG: TG mismatches. The $(GR)_{11}$ motif in MBD2 is a stretch of glycine argine residues and $(E)_{12}$ is a glutamate-rich domain. The Kaiso protein is characterized by three C2H2 zinc finger motifs (ZF) required for binding of methylated DNA (adapted from Clouaire and Stancheva, 2008).

MeCP2 is able to bind to a single methylated CpG pair via the 85 amino acid mCpG binding domain (Nan *et al.* 1993). However, it is thought that MeCP2 mediates gene repression through the TRD binding to the corepressor Sin3A (a paired amphipathic protein) which in turn recruits HDACs (Jones *et al.,* 1998). This causes changes in the chromatin architecture caused by the alteration of PTMs on the histone tails, condensation of the chromatin and subsequent inhibition of transcription. More recent research has shown that this is not the only model for MeCP2 mediated gene repression. Inhibition of MeCP2 mediated gene silencing by the addition of HDAC inhibitors such as trichostatin A (TSA) only partially relieves transcriptional repression, indicating that other pathways are involved such as blocking the access of transcription factors to methylated promoters it occupies or, alternatively, by

interacting directly with TFIID, a component of the basic transcriptional machinery (Wade, 2001).

Fuks *et al.,* (2003) showed through chromatin immunopreciptation (ChIP) experiments that MeCP2 is associated with histone methylation *in vitro* and *in vivo.* The methylation of H3K9 in the H19 gene is facilitated by MeCP2 and therefore it was proposed that MeCP2 might act as a 'bridge' between DNA and H3K9 methylation and acts to reinforce the repressive function of the two distinct methylation events.

MBP1 is the largest of the MBD family of proteins, consisting of 640 amino acid residues. The first two functional domains of the protein to be identified was the MBD and a region containing three CxxC- motifs (homologous to motifs found in DNMT1) and binds preferentially to densely methylated DNA *in vitro.* MBD1 has been shown to interact *in vivo* with the histone methyltransferase (HMT) SETDB1, which methylates H3K9 (Sarraf and Stancheva, 2004).

MBD2 and MBD 3 are structurally related, containing a similar exon-intron structure and have 70% similarity in their amino acid structures but differ in their functions. MBD2 forms the mCpG binding part of the MeCP1 complex. This complex is between 400 and 800kDa in size and can bind to twelve or more mCpGs in any sequence context. The MeCP1 mediated repression is different to the MeCP2 regulated repression as MeCP1 requires more densely methylated DNA for binding (Meehan, 1989). Unlike MBD2, MBD3 cannot specifically bind to methylated CpGs (Hendrich *et al.,* 1999). This is due to the substitution of a tyrosine residue which is

critical for activity to a phenylalanine residue (Fraga *et al.*, 2003). However, it has been found to be part of one of the major chromatin remodelling complexes, NuRD/Mi2 (nucleosome remodelling and histone deacetylase complex) (Wade *et al.*, 1999) and it is crucial for development as knockout of MBD3 leads to embryonic lethality in mice (Hendrich *et al.*, 2001).

The final MBD protein, MBD4, contains a glycosylase domain in addition to the MBD domain. MBD 4 co-localises with regions of constitutive heterochromatin. It has been shown to bind to methylated DNA *in vitro* but it does not play a role in the regulation of gene expression. It has a higher specificity for 5mCpG-TpG mismatched sites which are the result of the spontaneous deamination of methylated cytosine residues. MBD4 contains a carboxy terminal domain which is homologous to the bacterial DNA repair enzymes. This suggests that the protein is involved in the repair mechanism and functions to minimise mutations (Hendrich *et al.*, 1999).

1.5 CpG Islands

In the human genome approximately 80% of CpGs are methylated. However, there are some regions within the genome that have a significantly lower level of DNA methylation and these are called CpG islands (CGIs). They have a higher G + C content (67% compared to 41% in the rest of the genome, Antequera, 2003) but are hypomethylated. These islands vary in length from approximately 200bp to 3-4kb. The promoters of all housekeeping and 40% of tissue specific genes (56% of human genes) are embedded within a CGI (Antequera and Bird, 1993). However, CGIs can occur within genes and at intergenic regions and can be either methylated or methylation-free.

The significantly higher frequency of CpG dinucleotides compared to the expected number based on the DNA base composition is due to the inherent instability of methylated cytosine residues. Outside of CGIs, spontaneous deamination of methylated cytosine results in a thymidine residue, and generates a T:G mismatch in CpG dinucleotides. This will be fixed as TpG if not replaced by a cytosine before the next round of DNA replication. In contrast, deamination of an unmethylated cytosines yields uracil which results in a U:G mismatch. Subsequent control and repair mechanisms recognise uracil as an erroneous base and thus substitute it with cytosine and the CpG dinucleotide is re-formed. In contrast, the repair mechanisms for T:G mismatches involve different, less efficient mechanisms which results in the loss of cytosine residues after DNA replication (Antequera, 2003).

CGIs were originally identified in mouse genomic DNA. Digestion with the methylation sensitive restriction enzyme *Hpa*II (CCGG recognition site) which will only cut when the second cytosine residue is unmethylated resulted in certain regions of the genome being digested into very small fragments (termed *Hpa*II tiny fragments, or HTFs) and these were found to contain groups of non methylated CpG sites (Bird *et al.* 1985). These sequences were characterised and as a result the properties of CGIs were identified: a DNA sequence could be defined as a CGI when it was at least 200bp in length with a G + C content of >50% and a CpG frequency (observed/expected; o/e) of >0.6 (Gardiner-Garden and Frommer, 1987). More recently, the completion of the human genome project has allowed CGIs to be identified *in silico*. The computational prediction of CGIs using the original parameters led to a large number of putative sequences being identified which contained non CGI sequences, such as Alu elements. These are highly repetitive short

interspersed elements and have an approximate 280bp consensus sequence. In addition some can have a relatively high %GC and CpG o/e frequency. This led to the refinement of the criteria being used to define a CGI, with the minimum length of CGIs being increased to 500bp. This reduced the number of CGIs identified by approximately 90%, and contaminating Alu elements were mostly excluded (Takai and Jones, 2002). There is still some variation in the number predicted using these parameters depending on the algorithms used. However, most of the prediction and sequence selection techniques identifies between 24,000 and 27,000 CGIs in the human genome (Illingworth and Bird, 2009).

CGIs are inherently methylation free, although certain CGIs throughout the genome become methylated during development, which correlates with transcriptional repression. For example, CGIs on the inactive X-chromosome and those associated with imprinted genes all become heavily methylated. Therefore, CGIs are not immune to methylation. However, the mechanism causing CGIs to remain methylation free during de novo methylation in development remains unclear, although several hypotheses have been described. Figure 1.6 shows these potential mechanisms which result in the CGI hypomethylation.

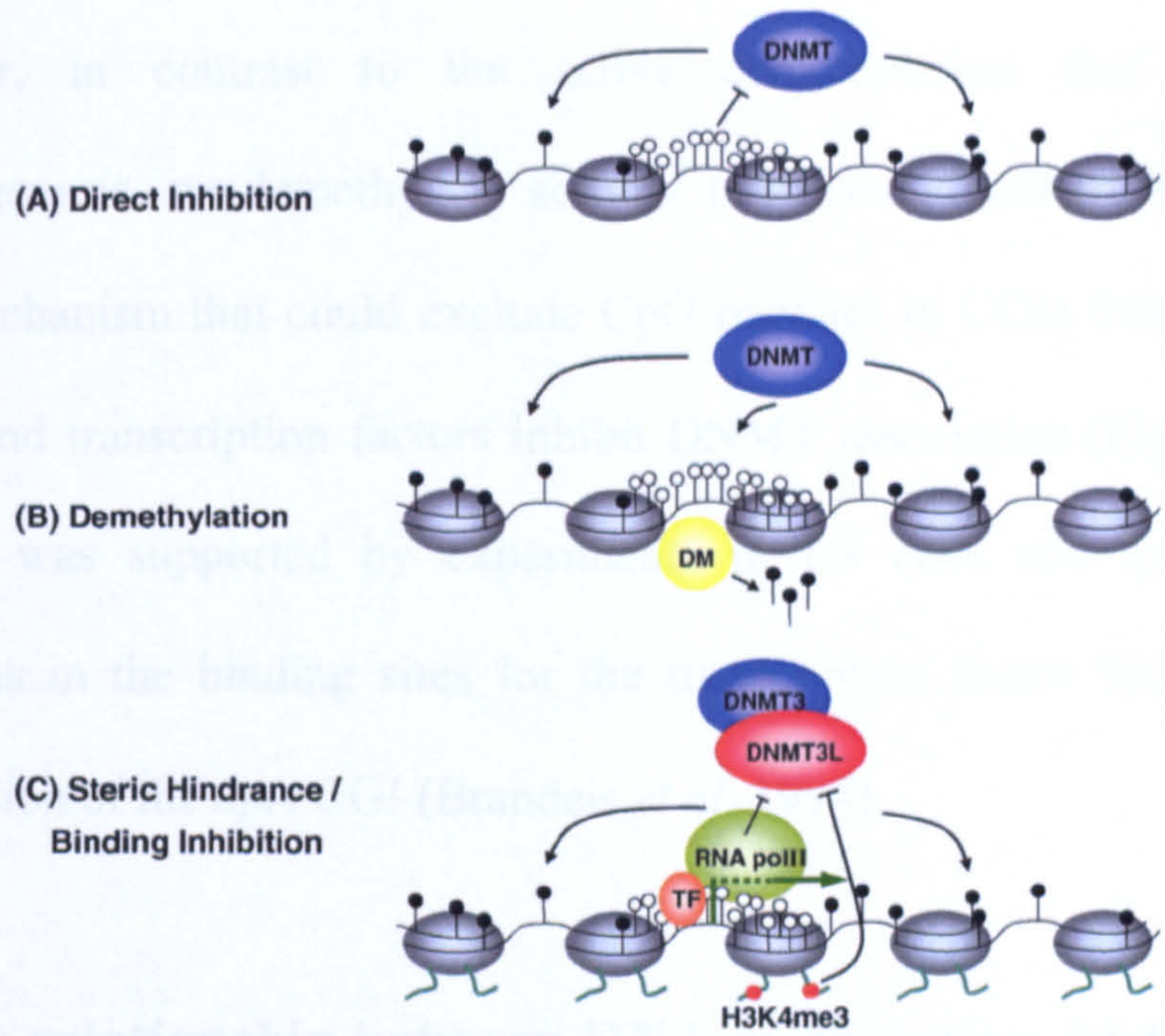


Figure 1.6 Potential mechanisms leading to CGI hypomethylation. A) CGIs remain hypomethylated due to intrinsic properties which stop the action of DNMTs. B) CGIs become methylated but are then demethylated by an island specific demethylating activity. C) The transcriptional machinery in addition to histone modifications (H3K4me3) stop the action of DNMTs from the transcriptional start sites. Methylated and unmethylated CpGs are denoted by filled and open lollipops respectively. (Illingworth and Bird, 2009)

The first hypothesis is that the CpGs in CGIs are resistant to methylation by DNMTs due to their DNA sequence (Figure 1.5A). However, this seems unlikely as the islands contain a higher frequency of the CpGs which are the preferred substrate of the DNMTs. Also CGIs within the inactive female X chromosome become methylated. A second hypothesis to account for CGIs remaining methylation free is that the islands are the target for demethylation where the methyl groups are actively removed from the cytosine residues (Figure 1.5B). This was illustrated in a study by Frank *et al.* (1991), where they investigated the fate of a methylated hamster adenine phosphoribosyltransferase (*aprt*) gene when introduced into fertilised mouse oocytes which were then implanted into recipient mice. Rapid demethylation of the 5' CpG

island region was observed but with the 3' region retaining its methylated status. However, in contrast to the active demethylation that takes place during embryogenesis, no demethylase activity has been identified in somatic cells. The third mechanism that could exclude CpG residues in CGIs from being methylated is that bound transcription factors inhibit DNMT association (Figure 1.5C). Evidence for this was supported by experiments in ES cells and transgenic mice where mutations in the binding sites for the transcription factor Sp1 resulted in *de novo* methylation of the *aprt* CGI (Brandeis *et al*, 1994).

1.6 The relationship between DNA methylation, histone modifications and chromatin remodelling

It is clear that the chemical modifications of DNA and histone complexes play a critical role in the regulation of gene expression. These modifications do not act in isolation, but instead are interconnected and cooperate in establishing and maintaining multiple layers of epigenetic gene regulation (Figure 1.7). DNA methylation and PTM of histone tails, in particular the deacetylation and methylation of certain residues, both result in gene silencing and although these modifications are performed by different sets of enzymes, and in different chemical reactions, there is clearly a biological relationship between these two systems.

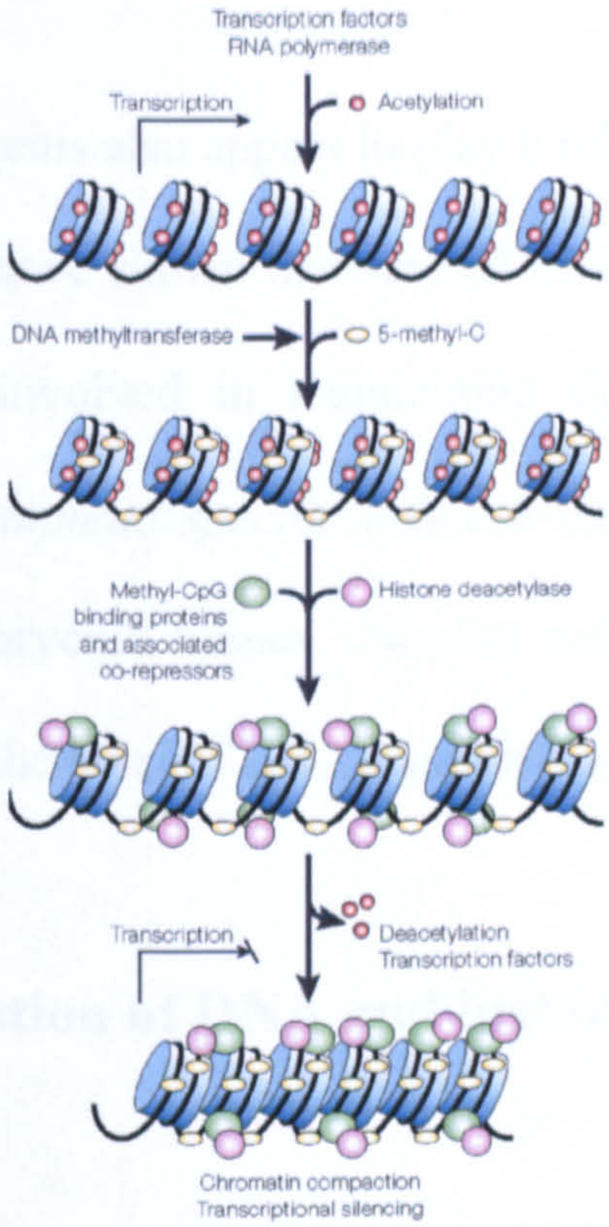


Figure 1.7 Cooperation of DNA methylation and histone deacetylation to repress transcription. A transcriptionally active region targeted for silencing is proposed to acquire DNA methylation first, which then recruits the methyl-CpG binding proteins and their associated co-repressors and histone deacetylases (HDACs). Taken from Robertson & Wolffe, 2000.

The precise sequence of events that leads to gene silencing is still unclear. However there is evidence that the relationship between DNA methylation and histone modifications is bidirectional: DNA methylation can influence histone modifications and in turn, histone modifications can influence DNA methylation. DNA methylation can influence histone modifications due to the DNA methylation machinery (DNMT1, MBD proteins) recruiting repressor complexes which contain HDACs (Nan *et al.*, 1998). In contrast, studies have shown that histone modifications can influence DNA methylation. It has been shown in fungi, plants and mammals that methylation of H3K9 can direct DNA methylation (Tamaru and Selker, 2001; Jackson *et al.*, 2002; Lehnertz *et al.*, 2003).

Chromatin remodeling proteins also appear to play a role in coordinated regulation of gene expression. Studies have shown that several of the SNF2 family of chromatin remodeling proteins are involved in maintaining the correct DNA methylation patterns. In mammals, *Lymphoid specific helicase* (*lsh*) is a SNF2 family member which is expressed in embryonic tissues. Studies using *lsh*$^{-/-}$mice showed that the mice had global defects in the level of DNA methylation (Dennis *et al.*, 2001).

1.7 Chemical modification of DNA and histone PTMs

A number of chemical entities that inhibit the repressive epigenetic effects of HDACs and DNMTs (iHDACs and iDNMTs) have been identified and gained FDA approval as therapeutic agents for a number of medical conditions including cancer (Baylin, 2005; Johnstone, 2002).

iHDACs can be divided into several families depending on their structural properties. These include hydroxamates, cyclic peptides, aliphatic acids and benzamines. These iHDACs are thought to inhibit the activity of class I and class II HDACs by chelating the zinc atom in the HDAC active site (Dokmanovic, 2007).

The inhibition of HDACs results in hyperacetylated histones and as a result nucleosomes are more relaxed and DNA is more accessible to regulatory factors, enabling efficient transcription. As a result iHDACs have also been used as enhancers of transgene expression in cultured mammalian cells. One of the most widely used is sodium butyrate, a member of the aliphatic acid family of iHDACs and

studies have shown that the addition of sodium butyrate to the culture medium increases product titres of recombinant proteins expressed in CHO cells (Mimura *et al.*, 2001; Palmero *et al.*, 1991; Jiang and Sharfstein, 2008). Although the addition of Sodium butyrate has been shown to be beneficial to transgene expression, it also has detrimental effects on the cell, including inhibition of cell growth and the induction of apoptosis (Davie, 2003).

iDNMTs include 5-azacytidine (5'AzaC) which can reactivate silenced genes caused by DNA methylation. This molecule was originally developed as a nucleoside anti-metabolite for acute myelogenous leukemia (Sorm *et al.*, 1964). It is a cytosine analog which becomes incorporated into newly synthesised DNA and forms covalent links with DNMTs (Jones and Taylor, 1980). This leads to the inhibition of DNMT activity in the cell and as a result cytosine residues become hypomethylated. The use of 5'AzaC and a derivative, 5-aza-2'-deoxycytidine, have been shown to partially restore transgene expression in cultured mammalian cells (Yang *et al.*, 2010; Kuriyama *et al.*, 1998)

The addition of both iHDACs and iDNMTs has been shown to have a synergistic effect, highlighting the link between histone acetylation and DNA demethylation in gene activation (Choi *et al.*, 2005).

1.8 Influencing chromatin structure to improve transgene expression

As has been described previously, recombinant protein expression levels are influenced by the chromatin structure surrounding the site of transgene integration

and/or the presence of regulatory elements. Therefore, the site of integration has a major effect on the transcription of a transgene. Integration into transcriptionally active regions of chromatin is preferential. However, in any given cell type the majority of mammalian genomes consist of heterochromatin and therefore the probability of the transgene being stably integrated into an area which is favourable for high level and stable expression is low (Grewal and Elgin, 2002). In addition, even if a transgene has integrated into a transcriptionally active region, expression may still be silenced as methylation of the DNA within the integrated transgene can occur regardless of the nature of the integration site including active regions of the genome (Pikaart *et* al., 1998; Yang *et al.*, 2010)

Investigators have found that the addition of genetic elements, which prevent neighbouring chromatin from affecting transgene expression is beneficial (Zahn-Zabal *et al.*, 2001). These reduce position effects and can also allow expression even if the gene has been integrated into an area of closed chromatin (Girod *et al.*, 2005). Elements that can potentially alleviate this problem have been identified in a variety of systems. These include insulators, Scaffold or Matrix Attachment Regions (S/MARs), Ubiquitous Chromatin Opening Elements (UCOEs) and Stabilising Anti Repressor (STAR) elements.

1.8.1 Insulators

Insulators are a class of DNA sequence elements that possess an ability to protect genes from the surrounding chromatin environment. They possess at least one of the two properties which define insulator function (Figure 1.8). Firstly, insulators can block the action of a distal enhancer on a promoter. This can only occur if the

insulator is placed between the enhancer and promoter. This activity limits the enhancer's function to only its target promoter. Secondly, insulators can reduce position effects by acting as a 'barrier' or 'border' (Sun and Elgin, 1999). For example, they can prevent the spreading of heterochromatin, which may otherwise silence gene expression.

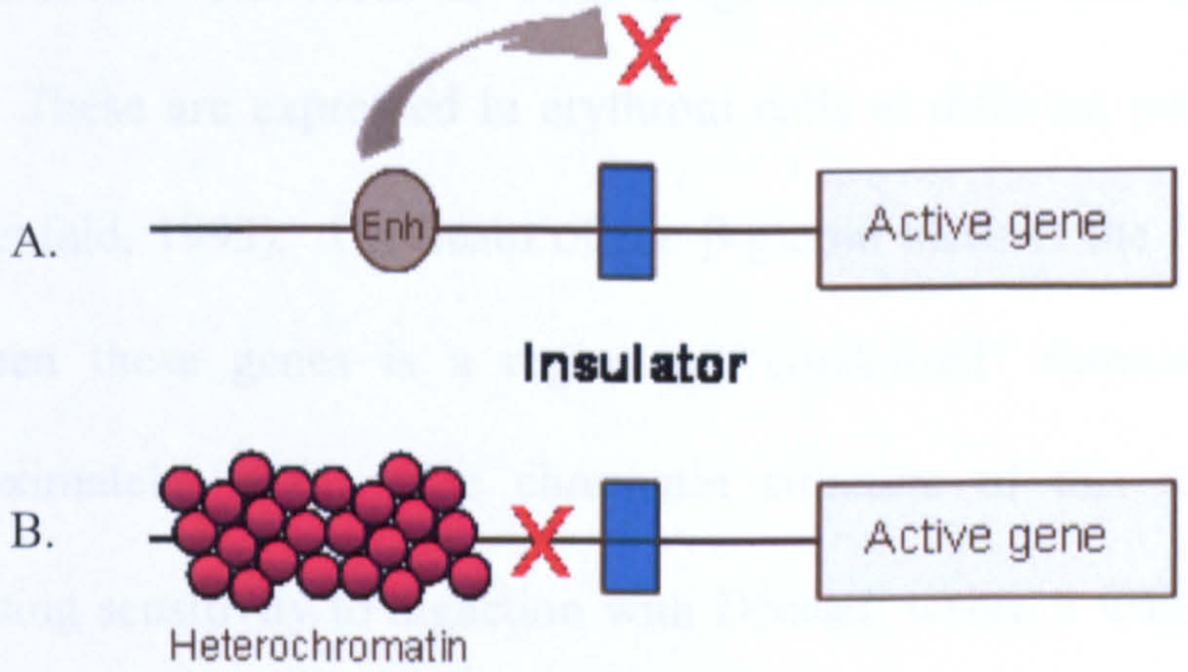


Figure 1.8 Insulator functions
A. Insulators can block an enhancer external to the active gene. B. A second mechanism of action is to block the spreading of heterochromatin and stop potential gene silencing (Adapted from Bell and Felsenfeld, 1999).

A large number of enhancer blocking elements have been identified, with insulators being described in *Drosophila* (Geyer and Corces, 1992), mammalian cells (Zhong and Krangel, 1997), chicken (Chung *et al.*, 1993) and *Xenopus* (Robinett *et al.*, 1997). One of the first examples of the enhancer blocking activity of insulator elements to be described was the *gypsy* retrotransposon from *Drosophila*. The insertion of this insulator between the 5' enhancer element and promoter of the *yellow* gene prevented the activation of the gene (Geyer and Corces, 1992). Another example from *Drosophila* are the *specialized chromatin structure* (*scs/scs'*) elements (Udvardy e*t al.* 1985). The *scs* and *scs*' elements are located downstream of *heat shock protein* 70 (*hsp*70) genes in the *Drosophila 87A7* heat shock locus and have been shown to block

repressive and activating chromosomal position effects from random sites of integration in *Drosophila* DNA (Kellum and Schedl, 1992).

1.8.1.1 Chicken HS4 (cHS4) insulator

In vertebrates, the first insulator to be identified was at the 5' end of the chicken β-globin locus. The locus is >33kb long, and includes four globin genes (see Figure 1.8). These are expressed in erythroid cells at different points during development (Felsenfeld, 1993). Upstream of the β-globin locus is the folate receptor gene, and between these genes is a region of 'condensed' chromatin, which extends for approximately 16kb. The chromatin structure of this region was analysed by assessing sensitivity to digection with DNaseI, where it was shown that the β-globin locus was more susceptible to digestion than the upstream condensed chromatin (Reitman and Felsenfeld, 1990). The mapping of DNaseI hypersensitive sites revealed constitutive elements that flanked the β-globin locus, 5'HS4 and 3'HS, with 5'HS4 forming part of the locus control region (Figure 1.9, top panel). In addition to ascertaining the chromatin structure using DNaseI, the histone modifications within this region were also analysed. The entire β-globin locus was found to be enriched for modifications associated with active transcription, including acetylation and methylation of lysine residues of histone H3. In contrast, within the condensed chromatin upstream of the locus these marks are depleted, and in their place there is a high percentage of methylation on the Lys-9 residue of histone H3, a mark of transcriptional repression (Litt *et al.*, 2001a; Litt *et al.*, 2001b). The 5'HS4 element (cHS4) marks the boundary of these two regions and it has been shown to have insulating functionality, due to its ability to confer position-independent expression of an eye colour reporter gene in *Drosophila* (Chung *et al.*, 1993). Investigation into the

insulating properties of the insulator showed that it had both enhancer blocking (Chung *et al.*, 1993) and barrier activity (Pikaart *et al.*, 1998). Characterisation of the insulator element identified a 'core' 250bp region which was sufficient for activity (Figure 1.9, bottom panel) (Chung *et al.*, 1997). Further investigation of this region by DNaseI footprinting revealed that the core element contains five protein binding sites and are referred to as footprints 1 to 5 (F1-F5, Figure 1.9). Functional analysis showed that only one of these (F2) was necessary for enhancer blocking (Bell *et al.*, 1999). Studies to identify proteins that bound to this region of DNA revealed that the protein CCCTC-binding factor (CTCF) was responsible for this activity. CTCF is an eleven zinc finger DNA-binding protein that is highly conserved in vertebrates and has been implicated in both transcriptional silencing (Filippova *et al.*, 1996) and activation (Vostrov and Quitschke, 1997). Functional analysis of the core region, employing a position-effect assay revealed that F2 and the CTCF binding site is not required for barrier activity and protection from position effects is mediated through the remaining four footprints (Recillas-Targa *et al.*, 2002). Chromatin position effect protection by 5'HS4 is associated with the enrichment of histone modifications near the insulator that are associated with transcriptional activation, including acetylation of histones H3 and H4 and methylation of H3K4 and H3R4. The enrichment of these histone modifications are mediated by proteins binding to the 5'HS4 insulator. The regulatory proteins Upstream Stimulatory Factor 1 and 2 (USF1 and USF2) bind to the insulator in F4 as a heterodimer and recruit HATs as well as the H3K4 methyltransferase SET 7/9 (West *et al.*, 2004) and the arginine methyltransferase PRMT1 (Huang *et al.*, 2007) to the area adjacent to the insulator. Although the histone modifications mediated by USF1/USF2 are essential for barrier activity, they are not sufficient to prevent gene silencing in vertebrates. Deletion of either F1, F3 or

F5 disrupts barrier activity without affecting USF-mediated recruitment of histone modifications. Previously it has been observed that transgenes lack promoter DNA methylation when shielded from chromosomal silencing by 5' HS4 (Mutskov *et al.*, 2002). Dickson *et al.* (2010) found that this hypomethylation was mediated by the binding of the Vascular Endothelial Zinc Finger 1 (VEZF1) transcription factor to footprints F1, F3 and F5.

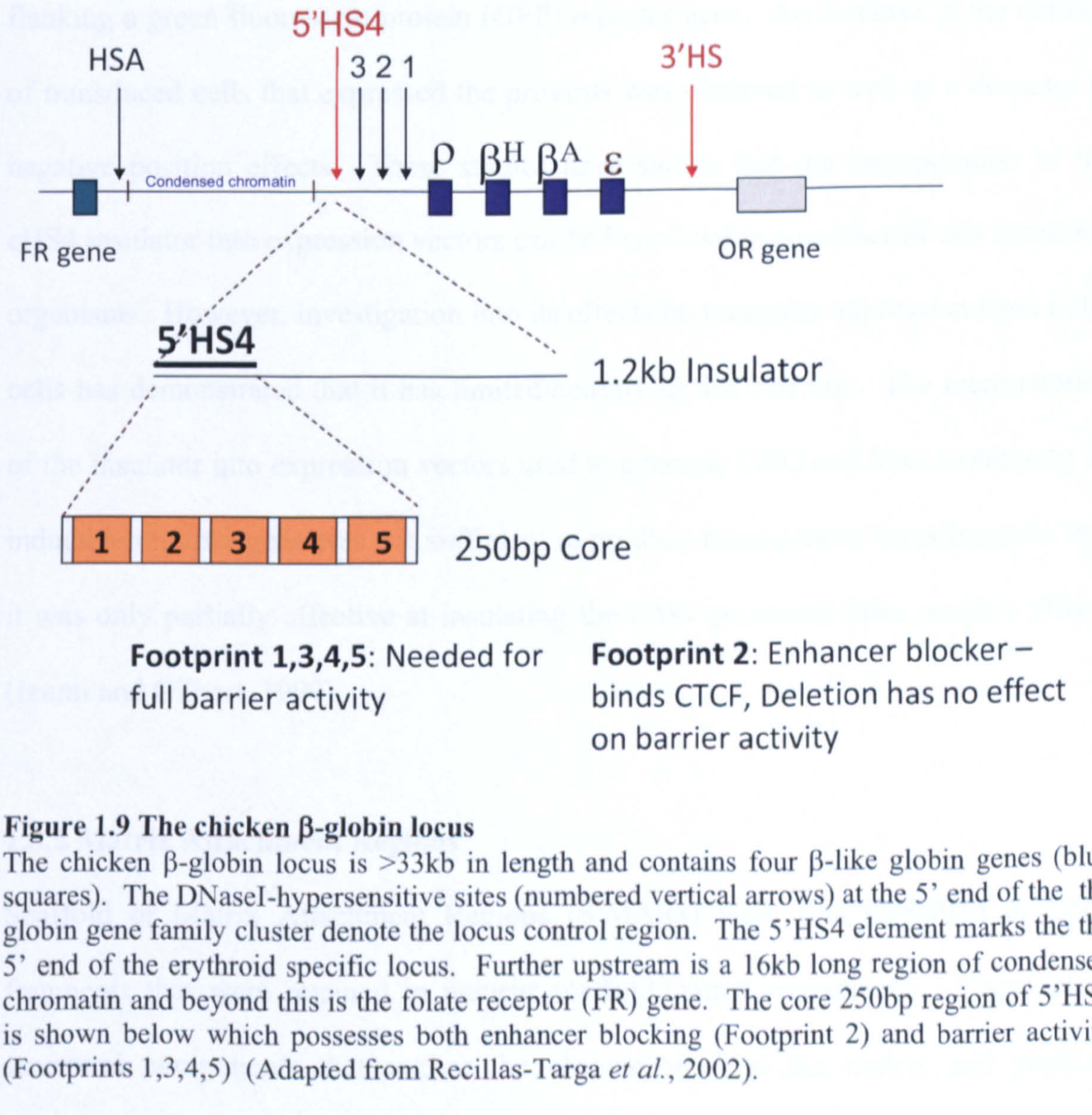


Figure 1.9 The chicken β-globin locus
The chicken β-globin locus is >33kb in length and contains four β-like globin genes (blue squares). The DNaseI-hypersensitive sites (numbered vertical arrows) at the 5' end of the the globin gene family cluster denote the locus control region. The 5'HS4 element marks the the 5' end of the erythroid specific locus. Further upstream is a 16kb long region of condensed chromatin and beyond this is the folate receptor (FR) gene. The core 250bp region of 5'HS4 is shown below which possesses both enhancer blocking (Footprint 2) and barrier activity (Footprints 1,3,4,5) (Adapted from Recillas-Targa *et al.*, 2002).

In addition to the work performed to characterise the cHS4 insulator where it was shown to protect against position effects in *Drosophila* (Chung *et al.*, 1993), its utility in other experimental settings has also been investigated. Studies have shown that the HS4 insulator can block promoter suppression in cultured human K562

erythroleukemia cells (Villemure *et al.*, 2001; Walters *et al.*, 1999). Functionality has also been demonstrated in the generation of transgenic mice. In a study by Potts *et al.* (2002), the expression of a tyrosinase minigene flanked by tandem copies of the cHS4 insulator resulted in a significant reduction in the variability between transgenic mice. Beneficial effects have been reported in gene therapy vectors. Emery *et al.* (2000) incorporated the cHS4 insulator into an oncoretroviral vector, with the insulator flanking a green fluorescent protein (GFP) reporter gene. An increase in the number of transduced cells that expressed the provirus was observed as well as a decrease in negative position effects. These studies have shown that the incorporation of the cHS4 insulator into expression vectors can be beneficial in a number of cell types and organisms. However, investigation into its effects on transgene expression from CHO cells has demonstrated that it has limited activity in this cell line. The incorporation of the insulator into expression vectors used to generate CHO cell lines expressing an inducible reporter gene was not sufficient to produce homogenous transformants, and it was only partially effective at insulating the CMV promoter from position effects (Izumi and Gilbert, 1999).

1.8.2 Matrix Attachment Regions

Scaffold or Matrix Attachment Regions (S/MARs) were first identified in DNA fragments that were retained in nuclear scaffold/matrix preparations. They act as structural components that anchor the chromosomes to the matrix and position chromatin into loop structures which can range in size from a few kb to more than 100 kb. In addition, they are thought to assist in the packaging of DNA in the nucleus (Bode *et al.*, 1996). The loop structures are formed because the DNA sequence of a MAR element consist of regions several hundred bp in length, which are extremely

AT rich. These irregularly spaced runs of A's and T's cause the DNA to become curved in a non-B DNA structure and MARs form the anchor points of the loop domains (Hart and Laemmli, 1998). MARs may also form genetic boundaries between chromosomal domains and as such only the encompassing *cis*-regulatory elements will control the expression of genes within the domain (Zahn-Zabal *et al.*, 2001). MARs have been shown to provide insulation from chromatin effects and also increase transcription initiation levels (Mielke *et al.*, 1990). Steif *at al.* (1989) identified the *cis* acting MAR regions flanking the chicken lysozyme gene (clysMAR), by detecting elevated expression of a reporter gene, which was independent of chromosomal region. Since this was reported there have been various studies showing that the clysMAR can protect against position effects and enhance gene expression in mammalian systems (Girod et al., 2005; Zahn-Zabal et al., 2001). Zahn-Zabel *et al.* (2001), showed that the presence of a clysMAR element in the expression vector resulted in an increase in the proportion of high expressing clones isolated as well as increasing the average expression levels of the clones analysed. This was achieved when the MAR elements were placed at both ends and thus flanked the transgene (in *cis*) as well as when constructs were co-transfected with the clysMAR on a separate plasmid (in *trans*). Girod *et al.*, (2005), performed further analyses investigating the effect of the clysMAR element on recombinant antibody production. In this study it was shown that expression levels were significantly increased when the clysMAR was added both in *cis* and *trans*. Other MAR elements have been shown to augment recombinant protein expression in mammalian cells including the MAR element from the human β-globin region. The β-globin locus is 20kb in length and the MAR element is located at the 5' boundary (Yu *et al.*, 1994). Kim *et al.* (2004) showed that the presence of this MAR element placed upstream of

the promoter driving β-galactosidase (β-gal) expression, could increase the frequency of positive clones by up to 80% and these colonies expressed β-gal approximately seven-fold higher than when no MAR was present. Wang *et al.* (2010) have also utilised the human β-globin MAR element and investigated its effects on the expression of a chloramphenicol acetyltransferase (CAT) reporter gene within CHO cells. Expression vectors were constructed with the MAR element at various positions around the CAT expression cassette. The configuration that resulted in the highest CAT expression levels was when two copies of the MAR element were present in the vector, flanking the CAT expression cassette.

Whilst there is an estimated 50,000 MAR elements in the human genome, only a few have been identified and characterised (Bode *et al.*, 2000). This is because MAR elements are not associated with any specific DNA sequence, and their function is more closely related to structural properties and characterised MAR elements had previously all been identified experimentally. With the aim of identifying novel MAR elements, Girod *et al.*, (2007), employed a computational approach to analyse putative MAR sequences in the human genome. Regions of the clysMAR that had been shown to increase transgene expression were used to scan the human genome for putative MAR sequences and approximately 54,000 potential MAR sequences were identified. By setting more stringent criteria to identify the most potent elements this figure was reduced to 1566 sequences. Cloning of several of these sequences and subsequent linkage and functional expression of a reporter gene, indicated that these MAR elements were more potent than the previously described clysMAR element.

1.8.3 Ubiquitously-acting Chromatin Opening Elements (UCOEs)

Certain gene products are essential for continuous cell growth and survival and the genes which encode these are referred to as housekeeping genes. The continuous expression of these genes requires the transcriptional machinery to have constant access to the DNA implying that these genes must reside in euchromatin. Harland *et al.* (2002) investigated the region of DNA surrounding a ubiquitously expressed housekeeping gene in order to identify any elements associated with this region of DNA. The study investigated the *TATA binding protein* (*TBP*) gene that is an essential component of the basal transcriptional machinery. Analysis of the TBP locus identified three functionally distinct genes, encoding TBP, Proteasome component-B1 (PSMB1) and Programmed cell death-2 (PDCD2). These genes are all housekeeping genes and possess a methylation free CGI spanning their promoter regions. Initially, Harland et al. showed that when murine fibroblast L-cells grown in tissue culture were stably transfected with the 44kb genomic region containing the entire *TBP* gene and the 5' region of the *PSMB1* gene, TBP could be constitutively expressed in all clones isolated at consistently high levels, and that chromatin position effects were negated. Further investigation into how an open functional chromatin domain is established at housekeeping gene loci was undertaken through the analysis of the organisationally similar heterogeneous nuclear ribonucleoprotein A2/B1 (HNRNPA2/B1)-heterochromatin protein 1Hs-γ (chromobox homolog 3, *CBX3*) locus (Antoniou et al., 2003). Stably transfected cells showed that transgenes containing methylation-free CGIs spanning the dual divergently transcribed promoters from both the *TBP-PSMB1* and the *HNRPA2B1-CBX3* gene loci were not silenced and no position effects were observed even when the transgene was integrated into centromeric heterochromatin. These regions of DNA were therefore named

Ubiquitous Chromatin Opening Elements (UCOEs). The pattern of DNA methylation and histone modifications across the *HNRPA2B1-CBX3* locus have been analysed (Lindahl Allen and Antoniou, 2007). A large region of the locus was found to be methylation free, which extended beyond the predicted CGI. In contrast, it was established that the 3' end of both the *HNRPA2B1* and *CBX3* promoters were methylated. However, analysis of the histone PTMs associated with these regions revealed that active acetylation and methylation marks on H3 and H4 were present. The ability of fragments of the *HNRPA2B1-CBX3* locus, when used in combination with a hCMV promoter, to increase and stabilise transgene expression in CHO-K1 cells has been evaluated (Williams *et al.* 2005). A series of three CGI-containing fragments of 8.0kb, 4.0kb and 1.5kb in length were investigated. The 8.0kb and 4.0kb fragments contained the complete CGI whereas the 1.5kb fragment contained only half of the CGI corresponding to the core region still spanning the dual divergently transcribed promoters of *HNRPA2B1-CBX3*. It was observed that vectors containing either of these three fragments gave at least a 20-fold increase in the level of enhanced green fluorescent protein (EGFP) produced in stably transfected pools of cells compared to that observed with the hCMV promoter alone. It was also shown that the level of expression remained stable for over 100 generations. The ability of the 8.0kb fragment to enhance clonal cell line production in CHO-K1 cells was also investigated. The presence of the 8.0kb fragment in the expression vector increased the number of high expressing clones produced and expression levels remained stable over time. Benton *et al.*, 2002, have also utilised the 8.0kb UCOE and shown that the addition of this element is beneficial to antibody expression, both in terms of the number of expressing clones isolated as well as increasing antibody expression levels. The ability of a UCOE to prevent DNA methylation of an integrated transgene has

also recently been reported (Zhang *et al.*, 2010). It was demonstrated that transgene expression in murine embryonic carcinoma P19 cells was stable when the UCOE was present in a self-inactivating lentiviral vector, and this was due to the transgene being resistant to DNA methylation.

Further research has been undertaken to identify new UCOEs. Novel UCOEs have been identified using microarray analysis of RNA from CHO-K1 and NS0 cells against known murine genes. Elements analysed were limited to sequences containing high CGI content and bidirectional promoters (Simpson *et al.*, 2006). DNA sequences from the 5' untranslated regions of ribosomal protein genes were shown to be capable of improving expression levels of operably-linked transcription units. One of these elements (3.2kb UCOE) has been used successfully to increase antibody expression from CHO stable cell lines (Ye *et al.*, 2010). In this study, the addition of the UCOE upstream of both the heavy and light chain promoters resulted in a significant improvement in productivity.

1.8.4 Stabilising Anti Repressor Elements

This class of element was identified using a genetic screen where only the cells transfected with a vector containing a DNA sequence which blocked chromatin associated repressors survived (Kwaks et al., 2003). A library of human genomic DNA fragments which were between 500bp and 2000bp in size were cloned between LexA binding sites and a *zeomycin* resistance gene in a mammalian expression vector. The LexA binding sites allowed the binding of targeted fusion proteins between the LexA protein and the polycomb group proteins HPC2 or HP1 which would repress gene expression unless blocked by an anti repressor element. The library was

transfected into U-2 OS cells and expression of the LexA fusion protein was induced. When zeomycin was added to the culture medium the majority of the clones were killed. The surviving colonies were screened with the HP1 and HPC2 fusion proteins to determine which colonies counteracted the repression. This resulted in the recovery of 65 plasmids which conferred survival. Further analysis of several of these DNA sequences, referred to as Stabilising Anti Repressor (STAR) elements, due to their ability to conteract chromatin associated repression, was performed in order to establish whether these elements could enhance recombinant protein expression in mammalian cells.

The generation of stable cell lines using vectors containing a STAR element (STAR 40) flanking the transgene resulted in more colonies expressing the gene compared to a control construct where no elements were present. It has also been shown that this increase in expression can be achieved in a number of cell lines and with different promoters. The elements are said to convey copy number dependent expression and be stable in the absence of selection pressure (Kwaks et al., 2003).

Further analyses of the identified STAR elements indicated that the STAR elements had variable activity in CHO cells and that a different STAR element, STAR 7, had better functionality than STAR 40 (Otte *et al.*, 2007). In the same study STAR 7 was evaluated in combination with STAR 67. The STAR 67 element was shown to have only low levels of anti repressor activity but enhanced protein expression when used in combination with other STAR elements (Otte *et al.*, 2007; Van Blokland *et al.*, 2007). Stable cell line generation using an expression vector incorporating both STAR 7 and 67 upstream of the EGFP cassette and STAR 7 downstream resulted in

colonies being isolated that expressed on average eleven-fold higher levels of EGFP than the control (Otte *et al.*, 2007). The effect of STAR elements on transgene expression has also been investigated in combination with the targeting of a specific HAT domain to a transgene promoter. A study by Kwaks *et al.*, (2005) demonstrated that the HAT domain of p300 was sufficient to increase expression levels of EGFP in stably tranfected clones but a decline in expression was observed when the cells were cultured in the absence of selection pressure. The addition of STAR elements to the expression construct increased expression levels and these remained stable when the cells were cultured in the absence of selection (Kwaks *et al.*, 2005).

1.9 Project aims

The aim of the study was to investigate whether the identified chromatin modifying elements (UCOE, MAR, STAR and cHS4) could increase the level and stability of antibody expression in stable cell lines in the mammalian expression systems used at UCB.

The elements chosen were incorporated into an antibody expression vector at different positions and combinations with the aim of determining which vector configuration was optimal for each element. This was achieved by analysing their effects on antibody expression in pooled and clonal CHO-K1 cell lines. The optimal vectors were then used in a comparative side-by-side study to further understand the effects of these chromatin modifying elements on antibody expression from stably transfected cells.

Work was performed to genetically characterise a number of clones harbouring constructs with each of the four chromatin modifying elements or the control vector with the aim of trying to obtain mechanistic insight into how these elements affected antibody expression.

The final aim of the thesis was to assess whether the presence of the chromatin modifying elements within the expression vectors could influence stability of the established clonal stable cell lines and to understand whether epigenetic regulation, in the form of DNA methylation was contributing to any loss in expression.

CHAPTER 2

Materials and Methods

2.1 Materials

2.1.1 Reagents and equipment

Chemical and biological reagents, unless otherwise stated, were obtained from Sigma-Aldrich (Poole, UK). Tissue culture media and components were obtained from Invitrogen (Paisley, UK). Plastic consumables were obtained from BD Falcon (Oxford, UK) except for ELISA plates which were Nunc™ brand, supplied by Fisher Scientific (Loughborough, UK).

Restriction enzymes, Taq DNA polymerase and Alkaline Phosphatase, Calf Intestinal (CIP) were obtained from Roche Diagnostics Ltd (Burgess Hill, UK) or New England Biolabs (Hitchin, UK). Antibodies were acquired from Jackson Immuno Research Inc, West Grove, PA, USA).

PCR was carried out using a T3000 Thermocycler from Biometra (Goettingen, Germany). Small scale centrifugation was carried out using an Eppendorf 5415D benchtop microfuge (Anachem Ltd. Luton, UK). Large scale centrifugation was carried out using a Sorvall Legend RT centrifuge (Thermo Scientific, Waltham, MA, USA). An ELx405 plate washer was supplied by Biotek (Biotek UK, Potton, UK). ELISA plates were read on a discovery HT-R plate reader from MWG-Biotech (Ebersberg, Germany).

2.1.2 Buffers and solutions

Buffer/ Solution	Composition
PBS	137mM NaCl, 2.7mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4
L-Broth	1% (w/v) Bactotryptone, 0.5% (w/v) Bacto-yeast extract, 17mM NaCl, pH7.4
S.O.C. medium	2% (w/v) Bactotryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10mM $MgCl_2$, 20mM glucose
Salt solution	1.2M NaCl, 0.06M $MgCl_2$
TAE	40mM Tris-acetate pH8 1mM EDTA
ELISA blocking buffer	PBS, 3% (w/v) BSA, 1% Tween20 (v/v)
ELISA conjugate buffer	PBS, 0.5% (w/v) BSA, 0.2% (v/v) Tween 20
ELISA wash buffer	PBS, 1% (v/v) Tween 20
TMB	0.1mg/ml 3,3,5,5'-tetramethylethyene benzidine 0.004% (v/v) hydrogen peroxide in 0.1M sodium acetate buffer, pH 6
20X SSC	3M NaCl, 0.3M sodium citrate
Depurination solution	250mM HCl
Denaturation solution	1.5M NaCl, 0.5M NaOH
Neutralisation solution	1.5M NaCl, 0.5M Tris-HCl,
Southern blot hybridisation buffer	DIG Easy Hyb (Roche Diagnostics Ltd.)
Low stringency buffer	2x SSC, 0.1% (v/v) SDS
High stringency buffer	0.5x SSC, 0.1% (v/v) SDS
Maleic acid buffer	0.1M Maleic Acid, 0.15M NaCl, adjusted to pH 7.5 with NaOH
Southern blot washing buffer	0.1M Maleic acid, 0.15M NaCl, pH 7.5, 0.6% (v/v) Tween 20
Southern blot blocking solution	Blocking reagent (Roche Diagnostics Ltd.) diluted to 1x with Maleic acid buffer
Detection buffer	0.1M Tris-HCl, 0.1M NaCl, pH 9.5
Antibody solution	Anti-Dioxigenin-AP diluted 1:10000 (75mU/ml) in blocking solution
Hypotonic solution	0.075M KCl
Fixative solution	Methanol:Acetic acid, 3:1 kept on ice
10x NT buffer	500 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 10 mM DTT, 0.5 mg/ml BSA)
FISH hybridisation buffer	50% (v/v) formamide, 10% (v/v) 20x SSC, 20% (v/v) dextran sulphate, 20% (v/v) ddH2O
Storage buffer	4x SSC, 0.05% (v/v) Igepal
FISH block buffer	Storage buffer, 0.25% (w/v) BSA

2.1.3 Antibodies

Antibody	Reactivity	Type	Use	Supplier
MOPC21	unknown	Monoclonal	ELISA Standard	UCB Discovery Research
goat anti-mouse Fc	Mouse Fc region	Polyclonal	ELISA coating	Jackson Immuno
Goat anti mouse Fc-HRP	Mouse Fc region	Polyclonal	ELISA detection	Jackson Immuno
anti-DIG-AP	Dioxigenin	Polyclonal	Probe detection Southern blot	Roche Diagnostics Ltd

2.1.4 Bacterial strains

MAX Efficiency® Stbl2™ Competent Cells (Invitrogen) were used for most transformations. For all other transformations XL1 blue supercompetent cells (Stratagene) or TOP10 cells (Invitrogen) were used.

2.1.5 Vectors

PAC clone RP4-736G20 was obtained from the Welcome Trust Sanger Institute.

pCR-Script was obtained from Stratagene.

PCR4-Topo was obtained from Invitrogen.

UCB antibody expression vectors are based on pEE12 (Bebbington, 1995) in which the backbone from *Sal*I to *Mlu*I has been replaced with that of pVAX (Invitrogen) to introduce *kanamycin* resistance and a pUC origin of replication to give high plasmid copy number in *E.coli* hosts.

The antibody expression vector used in this study (Ab535) consists of a light chain encoding Ab535 VL and murine kappa constant region and a heavy chain consisting of Ab535 VH and murine γ 1full length constant region. Transcription of the

immunoglobulin heavy and light chains are controlled by a modified version of the human immediate/early cytomegalovirus promoter and the SV40 polyadenylation signal sequence. A *kanamycin* resistance gene and pUC origin of replication allows replication and selection in bacteria.

2.1.6 Cell lines

The CHO-K1 cells used in this study are a suspension adapted variant grown in CD-CHO medium, derived at UCB from an original isolate obtained from the ATCC.

2.1.7 Chromatin modifying elements

The chromatin modifying elements used in this study along their sizes and source can be found in Table 1.

Element	Size (kb)	Source of element
UCOE	1.5	Licensed from Millipore
MAR X_S29	3.4	Selexis patent application WO 2005/040377
STAR 40	1	Kwaks *et al.* (2003)
STAR 7	2	plasmid kindly provided by Crucell
HS4 Tandem	0.5	Chung *et al.* (1997)

Table 2.1 Size and source of chromatin modifying elements used in study

2.1.8 Oligonucleotides

All oligonucleotides used can be found in Appendix 2.

2.2 Methods

2.2.1 PCR amplification of DNA fragments

Routine amplification of DNA was carried out using Taq polymersase. PCR reactions were carried out in a total volume of 50µl and contained approximately 50ng template DNA, 10pmoles of forward and reverse primer, 0.2mM dNTP mix (dATP, dCTP, dGTP, dTTP), 5µl 10X PCR buffer (100mM Tris-HCl, 15mM $MgCl_2$, 500mM KCl, pH 8.3), 1.25U Taq polymerase and sterile dH_2O. Reactions were cycled under the following conditions:

95°C	2 minutes	
95°C	20 seconds	30 cycles
55°C	20 seconds	
72°C	2 minutes	
72°C	5 minutes	

Following amplification, reactions were stored at 4°C.

When high GC content DNA was amplified the Taq polymerase was substituted for the Herculase II polymerase kit (Stratagene, La Jolla, CA, USA).

Post reaction, samples were analysed using agarose gel electrophoresis (See Section 2.2.2).

2.2.2 Agarose gel electrophoresis

Gel electrophoresis was carried out using 1% (w/v) agarose gel in TAE buffer containing 0.5µg/ml of ethidium bromide. Electrophoresis was typically carried out at 10V/cm and the DNA bands were visualised using a 254nm UV light. Tri-Dye DNA markers (New England Biolabs) were used to size DNA fragments. Electrophoresis was used both to analyse fragments following restriction enzyme digestion and for separation prior to purification of fragments used in subsequent manipulations. To

minimise DNA damage, illumination of DNA fragments for cloning was carried out using a 366nm UV light.

2.2.3 DNA fragment purification from an agarose gel

DNA fragments were purified from agarose gel using the QIAquick Gel Extraction Kit (Qiagen Ltd. Crawley, UK) according to the manufacturer's protocol. Briefly, the gel fragment was excised from the agarose gel, and weighed. 3 volumes of buffer QG was added to 1 volume of gel (100mg ~ 100µl). After incubation at 50°C for 10 minutes 1 gel volume of isopropanol was added to the sample and mixed. DNA was bound to a QIAquick column and centrifuged at 13,200rpm for 1 minute. The flow through was discarded and 500µl buffer QG was added. The sample was centrifuged for 1 minute at 13,200rpm and the flow through discarded. The column was washed by the addition of 750µl buffer PE. Following the removal of the flow through the column was centrifuged at 13,200rpm for a further 1 minute to remove any traces of ethanol from the column. The QIAquick column was placed into a clean microcentrifuge tube and to elute the DNA 50µl dH_2O was added to the centre of the QIAquick membrane and the column was centrifuged for 1 minute at 13,200rpm.

2.2.4 Restriction digestions of DNA

For restriction enzyme analysis 10 units of restriction enzyme was incubated with 500ng of DNA in the appropriate 10x buffer for 90 minutes at the appropriate temperature. When restriction digested DNA fragments were required for ligations 5µg of DNA was used in the reaction.

2.2.5 Phosphotase treatment

Vector fragments with compatible restriction enzyme ends had the 5' – phosphate groups removed by treatment with CIP to prevent self-ligation. Enzyme was added at 0.5 units per 1µg of DNA and incubated in the recommended buffer at 37°C for 15 minutes. To eliminate CIP from the reaction mix the DNA was resolved by agarose gel electrophoresis (see Section 2.2.2) and the DNA purified as in Section 2.2.3

2.2.6 Ligation of DNA fragments

DNA fragments were ligated into vector backbones using the Roche DNA Rapid Ligation Kit® (Roche Diagnostics Ltd.) according to the manufacturer's protocol.

Briefly, vector and insert DNA were added to 1x DNA dilution buffer to a final volume of 10µl. 10µl T4 DNA ligation buffer was added and mixed thoroughly. T4 ligase (1µl; 5 units per 1µl) was added, the contents mixed thoroughly and incubated for 5 minutes at 15-25°C.

2.2.7 E.coli Transformation

2.2.7.1 MAX efficiency Stbl2 E.coli transformation

An aliquot of the ligation mix (10µl) was added to 100µl MAX Efficiency Stbl2 competent cells (Invitrogen) according to the manufacturer's instructions. Briefly, after the addition of DNA to the competent cells the tubes were placed on ice for 30 minutes. Following a heat shock at 42°C for 25 seconds the tubes were returned to the ice for 2 minutes. S.O.C. medium (900µl) was added and the cells incubated whilst shaking (225rpm) at 30°C for 90 minutes. Cells were plated out onto L-broth agar plates containing the appropriate concentration of antibiotic and propagated overnight at 30°C.

2.2.7.2 XL1 Blue E.coli transformation

XL1 Blue competent cells (Stratagene) were transformed as in Section 2.2.7.1 but with the following alterations. An aliquot of ligation mix (10µl) was added to 50µl XL1Blue competent cells and after incubation on ice the cells were heat shocked for 45 second at 42°C and returned to ice for 2 minutes. S.O.C. medium (450µl) was added and the cells incubated whilst shaking (225rpm) for 60 minutes at 37°C. After plating out onto L-Broth containing the appropriate concentration of antibiotic the plates were incubated overnight at 37°C.

2.2.8 TOPO® cloning

Taq generated PCR products were cloned into the pCR4-TOPO vector (Invitrogen) according to the manufacturer's protocol. Briefly, 2µl of PCR product was added to a microcentrifuge tube. To this was then added 1µl of salt solution and 2µl of dH_2O, the contents mixed followed by addition of 1µl of TOPO® vector and incubation at 15-25°C for 5 minutes. Following ligation 2µl of ligation mix was used to transform TOP10 cells as described in Section 2.2.7.2.

2.2.9 Small scale plasmid DNA purification from transformed *E.coli*

Single colonies were picked and grown overnight in the appropriate volume of L-broth containing antibiotic selection. For small-scale plasmid DNA preparation, QIAprep spin miniprep kits (Qiagen Ltd.) were used according to the manufacturer's instructions. Briefly, overnight cultures (5ml) were centrifuged for 10 minutes at 3,000rpm and the bacterial cell pellet was resuspended in 250µl buffer P1. Buffer P2 (250µl) was added and the sample mixed. Buffer N3 (350µl) was added and the sample was mixed thoroughly. The sample was centrifuged for 10 minutes at

13,200rpm and the supernatant was transferred to a QIAprep spin column. After centrifugation for 30 seconds at 13,200rpm the column was washed with 750µl of buffer PE. The column was centrifuged for 30 seconds at 13,200rpm and the flow through discarded. The centrifugation step was repeated to remove any residual ethanol and the DNA was then eluted into 50µl dH_2O.

2.2.10 Large scale plasmid DNA purification from transformed *E.coli*

QIAfilter plasmid maxiprep kits (Qiagen Ltd.) were used to purify larger scale plasmid DNA in accordance with the manufacturer's instructions. Briefly, overnight cultures were centrifuged for 10 minutes at 3,000rpm and the bacterial pellet was resuspended in 10ml buffer P1, 10ml of buffer P2 was added and the sample mixed. After incubating at 15-25°C for 5 minutes, 10ml of chilled buffer P3 was added to the lysate and the sample mixed thoroughly. The lysate was poured into the barrel of a QIAfilter cartridge and incubated at 15-25°C for 10 minutes. A Qiagen-tip 500 was equilibrated by adding 10ml of buffer QBT and allowing the column to empty by gravity flow. The lysate was filtered and then added to the Qiagen-tip. After the lysate had entered the tip by gravity flow the column was washed twice with 30ml buffer QC. DNA was then eluted with the addition of 15ml buffer QF. DNA was precipitated by adding 10.5ml isopropanol and then centrifuged for 30 minutes at 3,500rpm. The pelleted DNA was washed with 70% (v/v) ethanol and then air-dried before being resuspended in an appropriate volume of dH_2O.

2.2.11 Genomic DNA extraction from CHO-K1 cells

High molecular weight gDNA was extracted from CHO-K1 cells using the Wizard® Genomic DNA Isolation Kit (Promega, Southampton, UK) according to the

manufacturer's instructions. Briefly, cells were pelleted by centrifugation at 13,200rpm for 15 seconds and 200µl PBS was added to wash the cells. The cell solution was centrifuged at 13,200rpm for 15 seconds and the supernatant removed. 600µl nuclei lysis solution was added and pipetted to lyse the cells. 3µl RNase A was added to the nuclear lysate and the sample mixed by invertion of the tube 2-5 times. The mixture was incubated for 15-30 minutes at 37°C and then allowed to cool to 15-25°C for 5 minutes. 200µl of protein precipitation solution was added and the sample vortexed at 13,200rpm for 20 seconds. The sample was then incubated on ice for 5 minutes. After centrifugation for 4 minutes at 13,200rpm the supernatant was transferred to a microcentrifuge tube containing 600µl isopropanol. The solution was mixed by gentle inversion until the white thread like strands of DNA formed a visible mass. To pellet the DNA the sample was centrifuged at 13,200rpm for 1 minute. The DNA was washed with the addition of 600µl 70% (v/v) ethanol and then pelleted by centrifugation at 13,200rpm for 1 minute. The supernatant was removed and the pellet was air-dried for 10-15minutes. DNA was resuspended in 100µl dH_2O.

2.2.12 RNA extraction from CHO-K1 cells

Total RNA was isolated from CHO-K1 cells using the RNeasy plus kit (Qiagen Ltd) according to the manufacturer's instructions. Briefly, cells were disrupted by adding 350µl RLT plus buffer (supplemented with 10µl β-mercaptoethanol per 1ml RLT plus buffer). The lysate was homogenised by adding the lysate to a QIAshredder spin column, and centrifuging for 2 minutes at 13,200rpm. The homogenised lysate was transferred to a gDNA eliminator spin column and the sample was centrifuged for 30 seconds at 10,000rpm. The column was discarded and 350µl of 70% (v/v) ethanol was added to the flow through and mixed well by pipetting. The sample was then

transferred to an RNeasy spin column. After centrifugation for 15 seconds at 10,000rpm the flow-through was discarded and 700µl buffer RW1 was added to the RNeasy spin column. The sample was centrifuged for 15 seconds at 10,000rpm and the flow-through discarded. 500µl buffer RPE was added to the column and the sample centrifuged for 15 seconds at 10,000rpm. A further 500µl buffer RPE was added to the column and the sample centrifuged for 2 minutes at 10,000rpm. To eliminate any possible carryover of buffer RPE, the RNeasy spin column was placed in a new 2ml collection tube and centrifuged for 1 minute at 13,200rpm. RNA was eluted by adding 30µl RNase-free water directly to the column membrane and the sample centrifuged for 1 minute at 10,000rpm. To increase the yield of RNA a further 30µl RNase-free water was added to the column membrane and the sample centrifuged for 1 minute at 10,000rpm.

2.2.13 Phenol-chloroform extraction and ethanol precipitation of DNA

An equal volume of Tris-buffered phenol was added to the DNA and vortexed for 10 seconds and then centrifuged for 1 minute at 13,200rpm. The top aqueous layer was decanted into a fresh microcentrifuge tube and an equal volume of chloroform added. The sample was vortexed for 10 seconds and then centrifuged for 1 minute at 13,200rpm. The upper aqueous layer was again transferred to a fresh microcentrifuge tube. The DNA was precipitated by the addition of 0.1x volumes of 3M sodium acetate pH 5.2 and 2.5x volumes 95% (v/v) pre-chilled (-20°C) ethanol, and mixed by invertion. The DNA was pelleted by centrifugation for 30 minutes at 14,000rpm and 4°C. The ethanol was decanted and the DNA pellet washed once with 70% (v/v) ethanol and centrifuged (14,000rpm for 10 minutes). The ethanol was decanted, the pellet air dried and resuspended in an appropriate volume of dH_2O.

2.2.14 DNA concentration and purity determination using optical density (OD) measurement

Plasmid DNA concentration was determined by measurement of the OD at 260nm using an Ultraspec 3100 pro spectrophotometer (Amersham Biosciences, Little Chalfont, UK) assuming an OD of 1 represents a concentration of 50µg/ml of double stranded DNA for a 1cm path length. The concentration of DNA was calculated using the formula shown below:

$$(OD_{260} \times 50 \times \text{dilution factor})/1000 = \mu g/\mu l$$

gDNA concentration was determined, as above, but measurement of the optical density at 260nm was performed using a Nanodrop 1000 (Thermo Scientific). High purity DNA had an OD260:280nm ratio of greater than 1.8.

2.2.15 RNA concentration and purity determination

Total RNA concentration was determined by measurement of the OD at 260nm using the Nanodrop 1000 (Thermo Scientific) assuming an OD of 1 represents a concentration of 40µg/ml of RNA for a 1cm path length. The concentration of RNA was calculated using the formula shown below:

$$(OD_{260} \times 40 \times \text{dilution factor})/1000 = mg/mL$$

High purity RNA had an OD260:280nm ratio of between 1.7-2.1.

2.2.16 DNA sequencing

DNA sequencing was performed in 0.2ml thin walled PCR tubes (Anachem Ltd, Luton, UK). Each reaction contained approximately 250ng DNA or 1µl of miniprep DNA, 3.2pmoles sequencing primer, 1µl sequencing buffer (400mM Tris, 10mM $MgCl_2$) and 1µl Big Dye Terminator Reaction Ready sequencing premix. Both

sequencing buffer and premix were supplied in the Big Dye Terminator Reaction Ready sequencing kit (Applied Biosystems, Warrington, UK). Distilled water was added to bring the volume to 7.5µl and the reaction was cycled under the following conditions:

96°C	30 seconds	
96°C	10 seconds	35 cycles
50°C	5 seconds	
60°C	4 minutes	

The DNA was then precipitated by the addition of 0.1 volumes 3M Sodium Acetate (pH 4.6) and 2.5 volumes of 95% (v/v) ethanol. Samples were incubated at 15-25°C for 15 minutes and then centrifuged at 1500rpm for 45 minutes at 4°C. The resultant pellets were washed with 70% (v/v) ethanol, air-dried and then resuspended in 10µl formamide.

The samples were then analysed on a 3100 automated DNA sequencer (Applied Biosystems). Data was analysed using Sequencher software (Gene Codes Corporation, Ann Arbor, MA, USA).

2.3 DNA Methods

2.3.1 Reverse transcription PCR for cDNA generation

mRNA was converted into cDNA using a High Capacity cDNA RT kit (Applied Biosystems) according to the manufacturer's protocol. Briefly, a mix of the following components was prepared on ice: 2µl 10x RT buffer, 0.8µl 25x dNTP mix (100mM), 2µl 10x RT random primers, 1µl multiscribe™ reverse transcriptase and 4.2µl

nuclease-free water. RNA (1µg) was diluted in nuclease-free water to a volume of 10µl. The reaction mixture and RNA sample were combined and incubated under the following conditions:

25°C 10 minutes
37°C 120 minutes
85°C 5 seconds

2.3.2 Quantitative PCR

TaqMan PCR assays were performed on gDNA samples for copy number determination or cDNA for mRNA level quantification in 384-well optical plates on an ABI Prism 7900HT Sequence Detection system (Applied Biosystems). Each sample was assayed in triplicate. A master mix was prepared which contained for each 10µl TaqMan reaction, 2µl DNA template, 2.83µl dH_2O, 0.17µl 60x TaqMan probe (consisting of forward and reverse primers, each with 1x concentration of 900nM and a 6-FAM™ dye-labelled TaqMan probe, 1x concentration, 250nM) and 5µl 2x Quantitect mix (containing HotStar® DNA polymerase, Quantitect probe PCR buffer, dNTP mix, ROX passive reference dye, 8mM $MgCl_2$) (Qiagen Ltd.). The samples were cycled under the following conditions:

50°C 2 minutes
95°C 10 minutes
95°C for 15 seconds } 40 cycles
60°C for 1 minutes }

The results were analysed using the SDS 2.3 software for absolute quantification calculations or RQ manager software for relative quantity determinations. (Applied Biosystems).

2.3.2.1 Generation of standard curves for copy number determination

Standard curves were required for copy number determination. Plasmid DNA was used in the range of $1x10^6$ copies to $4.15x10^3$ copies in a 1:3 dilution series in CHO-K1 gDNA. Plasmid DNA molecular weight was calculated using the following:

If each bp of DNA = 660Da, and 1Da = $1.66x10^{-24}$g

Plasmid size (Ab535GS) = 11560bp

Plasmid M_r = 11560 x 660 = 7629600Da

7629600 x $1.66x10^{-24}$ = $1.27x10^{-17}$g

So, 1 copy of Ab535GS plasmid = $1.27x10^{-17}$g

and $1x10^8$ copies = $1.27x10^{-17}$ x $1x10^8$ = $1.27x10^{-9}$g = 1.27ng

using calculation above, M_r of other plasmids:

plasmid	size (bp)	ng DNA for $1x10^8$ copies
1.5kb A2UCOE	16236	1.78
MAR X_S29	18242	2.00
STAR 7	15849	1.74
cHS4 tandem	12941	1.42

2.3.3 Southern Blot Hybridisation

2.3.3.1 DIG labelled probe synthesis.

PCR was used to generate the DIG labelled probe used in the hybridizations.

PCR reactions were carried out using the Expand High Fidelity PCR system (Roche Diagnostics Ltd) in a total volume of 50µl and contained approximately 10pg template DNA, 10pmoles of forward and reverse primer, dNTP mix (200µM dATP, 200µM dCTP, 200µM dGTP, 130µM dTTp, 70µM DIG-11-dUTP), 10µl 10x buffer

with MgCl2, 2.6U of Expand High Fidelity enzyme mix and sterile dH_2O. Reactions were cycled under the following conditions:

95°C	2 minutes	
95°C	30 seconds	} 30 cycles
60°C	30 seconds	
72°C	40 seconds	
72°C	7 minutes	

PCR products were analysed using agarose gel electrophoresis (as described in Section 2.2.4).

2.3.3.2 DNA extraction and electrophoresis

Total gDNA (10µg) was digested to completion using *Eco*RI restriction enzyme for 16 hours. Digested DNA, along with DIG labelled DNA Molecular weight marker (Roche Diagnostics Ltd) was separated according to size by electrophoresis through a 0.8% (w/v) TAE agarose gel. After visualisation of the gel to check separation of the DNA the gel was placed in depurination solution for 15 minutes at 15-20°C. The gel was rinsed in de-ionised water and then incubated in denaturation buffer for 15 minutes in order to denature the double stranded genomic DNA. The denaturation buffer was replaced and the gel incubated for a further 15 minutes. It was then placed in neutralisation buffer for 15 minutes with shaking. The neutralisation buffer was replaced and the gel incubated for a further 15 minutes. After rinsing in dH_2O, the gel was equilibrated in 20x SSC for 10 minutes.

2.3.3.3 Capillary blot and DNA cross linking

A capillary blot was set up according to the Roche DIG Application Manual for Filter Hybridisation. Briefly, a piece of Whatman 3mm paper (GE Healthcare, Little Chalfont, UK) soaked in 20x SSC was placed on top of a 'bridge' so that it was resting in a shallow reservoir of 20x SSC. The gel was placed upside down on top of

the wet sheet of 3mm Whatman paper. A piece of positively charged nylon membrane was cut to the same size of the gel and placed on top. The blot assembly was completed by adding a dry sheet of 3mm Whatman paper, a stack of paper towels, a glass plate and a 200g weight. DNA was transferred overnight. Following transfer, the membrane was washed briefly in 2x SSC and then the DNA was fixed to the membrane by incubation in a Techne Hybridiser HB1D hybridisation oven for 2 hours at 80°C (Techne Inc, Burlington, NJ, USA).

2.3.3.4 Pre-hybridisation of membrane with DIG Easy Hyb

The correct hybridisation temperature for hybridisation in DIG Easy Hyb buffer was determined using the following calculation, as described in the Roche DIG Application Manual for Filter Hybridisation:

$T_m = 49.82 + 0.41$ (% G + C) – 600/l

$T_{hyb} = T_m$ – (20° to 25°C)

Where:

T_m = melting point of probe-target hybrid

(% G + C) = % of G and C residues in probe sequence

T_{hyb} = Optimal temperature for hybridisation of probe to target in DIG Easy Hyb

l = length of hybrid in base pairs

The probe being used had a G+C content of 46.9% and a length of 488 base pairs

Therefore:

$T_m = 49.82 + 0.41\ (46.9) - 600/488$

$49.82 + 19.229 - 1.23$

67.8

$T_{hyb} = 67.8 - (20°\text{ to }25°C)$

$= 42.8 \text{ to } 47.8°C$

A hybridisation temperature of 45°C was chosen.

The membrane was added to a hybridisation bottle along with 10ml of DIG Easy Hyb. The membrane was incubated at 45°C for 1 hour in a Techne Hybridiser HB1D hybridisation oven (Techne Inc.).

2.3.3.5 Hybridisation of DIG-labelled DNA probe to DNA on membrane

DIG labelled probe (20µl), (as synthesised in Section 2.3.1) was placed in a microcentrifuge tube containing 30µl dH_2O. The microcentrifuge tube was incubated at 100°C for 5 minutes to denature the probe and then chilled quickly in an ice bath. Prehybridisation buffer was removed and the labelled probe added along with the pre-warmed (45°C) DIG Easy Hyb buffer. The membrane was incubated with the probe overnight at 45°C.

2.3.3.6 Washing the DIG labelled membrane

After discarding the hybridisation mixture, the membrane was washed twice in 200ml low stringency buffer for 5 minutes at 15-10°C. The low stringency buffer was poured off and immediately replaced with the pre-warmed (65°C) high stringency buffer and incubated for 20 minutes at 65°C.

2.3.3.7 Detection of DIG labelled probes

The membrane was rinsed briefly for 5 minutes in washing buffer and then transferred to blocking solution where it was incubated whilst shaking at 15-25°C for 1 hour. The blocking solution was replaced with the anti-DIG antibody solution and incubated for 30 minutes. The membrane was washed twice in washing buffer for 30 minutes. After washing, the membrane was allowed to equilibrate in detection buffer for 5 minutes. For chemiluminescent detection of the hybridised probe the membrane was placed with the DNA side facing up on a plastic folder and 1ml of the alkaline phosphatase substrate CSPD added. A second sheet of plastic was placed immediately over the membrane and incubated for 5 minutes. The edges of the plastic folder were sealed and the folder was incubated for 10 minutes at 37°C. The membrane was exposed to X-ray film (Kodak, Hemel Hempstead, UK) for at least 30 minutes and the film developed using a compact X4 automatic X-ray film processor (Xograph Healthcare Ltd, Tetbury, UK)

2.3.5 Bisulphite conversion and sequencing

The complete conversion of unmethylated cytosines to uracils in 1µg gDNA was carried out using the EpiTect Bisulphite kit (Qiagen Ltd) according to the manufacturer's instructions. Briefly, the bisulphite mix was dissolved by adding 800µl RNase-free water. gDNA (1µg) was diluted in a total volume of 20µl RNase-free water. To this 85µl of the dissolved bisulphite mix was added. Finally, 35µl DNA protect buffer was added, the contents mixed thoroughly and the tube incubated in a thermal cycler under the following conditions:

95°C 5 minutes
60°C 25 minutes
95°C 5 minutes
60°C 85 minutes
95°C 5 minutes
60°C 175 minutes

The sample was transferred to a fresh 1.5ml microcentrifuge tube and 500µl buffer BL was added. The solutions were mixed by vortexing and then transferred to a EpiTect spin column. The spin column was centrifuged for 1 minute at 13,200rpm and the flow-through discarded. Buffer BW (500µl) was added to the spin column and the sample centrifuged for 1 minute at 13,200rpm. After discarding the flow-through 500µl buffer BD was added to the spin column and incubated for 15 minutes. The spin column was centrifuged for 1 minute at 13,200rpm and the flow-through discarded. Buffer BW (500µl) was added and the sample was centrifuged for 1 minute at 13,200rpm. After repeating the buffer BW wash the spin column was placed in a new 2ml collection tube and centrifuged for 1 minute at 13,200rpm to remove any traces of buffer BW. The spin column was then placed in a new 1.5ml microcentrifuge tube and 20µl buffer EB was pipetted onto the centre of the membrane. The DNA was eluted by centrifugation for 1 minute at 12,000rpm. To increase the yield of DNA a further 20µl of buffer EB was added to the membrane and the column centrifuged for 1 minute at 13,200rpm.

Following the bisulphite conversion and purification of the DNA a first round of PCR reactions were conducted. Reactions were set up as described in Section 2.2.1 using Taq polymerase but with the following alterations. The reaction mix contained 8µl of bisulphite converted DNA and the reactions were cycled under the following conditions:

94°C 2 minutes
94°C 30 seconds
55°C 30 seconds } 40 cycles
64°C 45 seconds
64°C 7 minutes
4°C Constant

An 8µl aliquot of the PCR reaction from the first round of reaction was used in a nested PCR reaction. Reactions were set up and cycled under the same conditions as the first round PCR.

2.4 Cell Culture Methods

2.4.1 Cell maintenance

Suspension adapted CHO-K1 cells were grown in CD-CHO media supplemented with 8mM Glutamax.

For routine passage cells were seeded at $2x10^5$ cells/ml every 3 – 4 days and grown shaking at 125rpm at 37°C in 8% CO_2. In addition, stable cell lines were also grown in 6 well plates and were passaged by diluting 1:10 into fresh CD-CHO media every 3-4 days.

2.4.2 Cell number determination

Cell number and viability were determined using a CEDEX A520 cell counter (Roche Innovatis AG, Bielefeld, Germany).

2.4.3 Pooled stable cell line generation

2.4.3.1 Neomycin selection

For pooled stable cell lines generated under Neomycin selection CHO-K1 cells were transfected using Lipofectamine 2000. A total of $1x10^7$ cells were transfected according to the manufacturer's protocol. Briefly, 15µg of circular DNA was diluted in 375µl of OPTI-MEM® I reduced serum medium in a sterile tube. Lipofectamine 2000 reagent (37.5µl) was diluted in another 375µl aliquot of OPTI-MEM® I medium and incubated for 5 minutes at 15-25°C. The Lipofectamine 2000 reagent and the DNA solutions were then combined and mixed gently. The mixture was incubated at 15-25°C for 20 minutes to allow DNA-reagent complexes to form. CHO-K1 cells were resuspended in 3.75 ml of fresh CD-CHO medium in a T75 flask. After the 20 min incubation, the reagent-DNA complexes were added to the T75 flask and incubated with the cells at 37°C for 4-4.5 hours. After the incubation, 6.75ml of fresh CD-CHO medium was added to the flask, and incubated overnight at 37°C. Stably transfected cells were then selected for by the addition of 1.5mg/ml geneticin sulphate (G418). FCS 1% (v/v) (PAA Laboratories Ltd, Yeovil, UK) was added to allow surviving cells to adhere and the medium was changed after one week. Two weeks post transfection when cell number and viability had recovered, cells were transferred to shaking culture and overgrow cultures set up to assess antibody expression levels.

2.4.3.2 Glutamine synthetase selection

Stably transfected pools of CHO-K1 cells under GS selection were transfected by electroporation as follows. Cells (1 x107) were resuspended in CD-CHO medium (700µl) and 100µl of linearised DNA (at a concentration of 400µg/ml) were mixed together in a Bio-Rad 0.4cm electroporation cuvette. The cell/DNA mixture was

electroporated with a single pulse of 300 volts and 960μF using a Bio-Rad gene pulsar unit (Bio-Rad Laboratories Ltd, Hemel Hempstead, UK). The cell suspension was carefully removed from the cuvette and placed into 50ml of warm CD-CHO medium. The cells were incubated overnight at 37°C in a static T175 flask. Stable transfectants were then selected by addition of 50μM MSX to the media 24 hours post transfection. After 10 days the cell suspension was centrifuged at 1,000rpm and the cells were resuspended in fresh CD-CHO medium supplemented with 50μM MSX. Once cell viability and number had recovered, overgrow cultures were set up to assess antibody expression levels.

2.4.4 Clonal Stable Cell Line Generation

2.4.4.1 Neomycin selection

Clonal cell lines were isolated by plating pools (as generated in Section 2.4.3.1) in CloneMedia (Genetix, New Milton, UK). For each cell line 12000 cells were added to 24ml of CloneMedia supplemented with 8mM Glutamax. The mix was plated out at 2ml/well (500 cells/ml) into 6 well equiglass cell culture plates (Genetix). Clones were picked using a Clonepix^{FL} robot (Genetix) under white light illumination into 96 well plates containing XP media (Genetix) supplemented with 8mM glutamine and 1.5mg/ml G418.

2.4.4.2 Glutamine synthetase selection

Clonal stable cell lines were transfected as described in Section 2.4.3.2. At 24 hours post transfection, cells were plated into 96 well plates at a cell concentration of $1x10^3$ cells per well in CD-CHO supplemented with 50μM MSX. Wells were scored manually for clonal cell growth 3 to 4 weeks later.

2.4.5 Determining specific antibody productivity

2.4.5.1 The integral of viable cell concentration

The integral of viable cell (IVC; 10^6 cells.h/ml) can be represented by the area under the cell growth curve. This integral was calculated by summing the areas under the curve between successive determinations of cell concentration. The method is based on that described by Renard *et al.* (Renard *et al.*, 1988) and can be calculated using the formula below

$$IVC = \frac{Xv_0 + Xv_1}{2} . (t_1 - t_0)$$

Where:

Xv_0	Viable cell concentration at first sample (10^6 cells/ml)
Xv_1	Viable cell concentration at second sample (10^6 cells/ml)
t_0	Elapsed time at first sample (hours)
t_1	Elapsed time at second sample (hours)

2.4.5.2 Specific rate of product synthesis

Values for the specific rate of antibody synthesis, Q_P (mg/10^6 cells/h), were calculated by a single point calculation. Using this method the specific rate of product synthesis is calculated by dividing the value for product concentration at harvest by the value for the time integral of viable cell concentration at harvest.

2.5 Assay Methods

2.5.1 Mouse IgG ELISA

Supernatant from the culture flasks were removed and antibody titre determined using a mouse IgG ELISA. Nunc™ 96 well plates were coated with 100µl of 2µg/ml goat anti-mouse Fc antibody in PBS overnight at 4°C. After washing twice with ELISA

wash buffer plates were blocked with 200µl/well ELISA blocking buffer for 1 hour at 15-25°C. Plates were washed as before and then samples were added to the top row of the plate. In addition a standard, MOPC21, (at a concentration of 1µg/ml) was added in duplicate and samples were titrated down the plate in ELISA conjugate buffer with tripling dilutions. After 1 hour incubation the plates were washed again and 100µl/well anti mouse Fc-HRP at a 1:5000 dilution in conjugate buffer was added to all wells and incubated at 15-20°C for 1 hour. Plates were washed again and binding revealed using TMB (Calbiochem, San Diego, CA, USA). Plates were read at 630nm and concentrations calculated using KC4 data analysis software (Biotek UK).

2.5.2 Spot ELISA

When large numbers of samples were assayed, a modified mouse IgG ELISA was conducted. One well per supernatant sample was utilised with no standard. Absorbance at 630nm was measured.

2.6 Cytogenic Methods

2.6.1 Fluorescence In Situ Hybridisation

2.6.1.1 Probe generation using nick translation

A reaction mix was prepared containing 1x NT buffer, 10mM β-mercaptoethanol, 0.5mM each of dATP, dCTP, 0.1mM dTTP, and 1mM biotin labelled dUTP (Roche Diagnostics Ltd.), 5ng DNase and 10U *E.coli* polymerase I (New England Biolabs). DNA template (1µg) was added in addition to DNase/RNase free water in a total volume of 50µl. The reaction was incubated at 16°C for 90 minutes. The reaction was stopped by adding 5µl of 0.5M EDTA, pH 8. To determine whether the probe

had been synthesised into the correct size (<500bp) a fraction of the reaction was visualised on a 1% (w/v) TAE agarose gel.

2.6.1.2 Preparation of metaphase chromosome spreads from cultured cells

Demecolcine (0.05µg/ml) solution was added to a 10ml cell culture for 1 hour to accumulate cells in metaphase. The cell medium and cells were transferred to a 15ml Falcon tube and centrifuged at 800rpm for 5 minutes at 15-25°C. Supernatant was discarded and 6ml of hypotonic solution, pre-warmed to 37°C, was added drop wise while resuspending the cells by gently flicking the tube. For the remainder of the 15 minutes, tubes were incubated in a water bath at 37°C. Tubes were centrifuged at 800rpm for 5 minutes at 15-25°C and the supernatant discarded. Cells were fixed by the addition of 5ml fixative, added drop wise to avoid cell clumping. Cells were resuspended by gently flicking the tube. Tubes were centrifuged at 800rpm for 5 minutes at 15-20°C and the supernatant was discarded. The fixing steps were repeated a further 3 times. Fixed cell suspensions can be stored long-term in fixative at -20°C.

2.6.1.3 Dropping slides

Fresh fixative was prepared and kept on ice. Slides were cleaned by incubating in methanol for several minutes at 15-20°C and then dried. The supernatant from the fixed cell suspension (prepared in Section 2.6.1.2) was removed until about 0.5ml of fixative was left and the cells were resuspended. Depending on the size of the pellet a further 0.5 to 2ml of fixative was added. A fraction of the cell suspension (8µl) was pipetted onto 3 separate areas of the same slide. Once the drops had a grainy appearance a further drop of fixative was added to the suspension. The suspension

was air dried and then the slides inspected under a phase contrast microscope to confirm metaphase spreads. Slides can be stored in 70% EtOH at 4°C at least for 6 weeks or dry at -70°C (in containers sealed in plastic bags containing desiccative) for at least one year.

2.6.1.4 Aging

The dropped slides were then aged by dehydrating them in an ethanol series (70%, 80%, 99.7%, all v/v) for 3 minutes each and then air dried. Subsequently the slides were incubated at 75° for 1 hour.

2.6.1.5 RNase treatment

Following aging, RNase treatment was performed. The RNase mix was prepared (2.5µl RNase, 97.5µl 2x SSC per slide) and vortexed briefly and then added to the slide. A 22x50mm cover slip was added and the slide incubated in a wet chamber at 37°C for 1 hour.

2.6.1.6 Preparation and denaturation of probes

Probe DNA (1µl) (as generated as in Section 2.6.1.1) and competitor DNA (such as herring sperm DNA, 10µg/µl) was added to a microcentrifuge tube. A 2µl aliquot of FISH hybridisation buffer was added and after mixing briefly the tube was centrifuged. The probes were denatured at 75°C for 5 minutes.

To denature chromosomes, the slides were first dehydrated in an ethanol series (70%, 80%, 100%) for 3 minutes each and then air dried. The slides were then placed at 75°C in 70% formamide (v/v) in 2x SSC for 90 seconds. The slides were then

dehydrated in an ethanol series (70% ice-cold, 80% RT, 100% RT, 3 minutes each), slides can be stored overnight if neccessary.

2.6.1.7 Hybridisation

The denatured probe was then added to the slides and covered with a 13mm cover slip. The slide was sealed with rubber cement. The slides were incubated over night at 37°C in a humidified chamber.

2.6.1.8 Post hybridisation washes

Coplin jars were arranged on the bench in the following order (1) 2x SSC; (2) 2x SSC, 0.1% Igepal; (3) storage buffer; (4) FISH block buffer.

The rubber cement was removed and the slides were washed in 2x SSC to remove the coverslips. The slides were then washed twice in 50% formamide (v/v) in 2X SSC at 37°C for 10 minutes each time. Slides were put into 0.1% Igepal in 2x SSC, for 1 minute at 15-20°C and then placed in storage buffer for 15 minutes. Slides were finally put in FISH block buffer for 25 minutes at 15-20°C.

2.6.1.9 Detection

After incubation in FISH block buffer probes were detected. All subsequent steps were performed in the dark. Detection mix was prepared (per slide 0.5µl of 5µg/ml streptavidin-Cy3 + 50µl storage buffer + 50µl FISH block buffer) and added to each slide and a cover slip placed on top. The slide was incubated at 37°C in a humidified chamber for 35 minutes. The cover slip was removed and the slides washed in storage buffer 3 times each for 3 minutes on a shaker. The slides were rinsed twice in dH_2O and air-dried.

2.6.1.10 Counterstaining with DAPI

Two drops of vectashield containing DAPI (Vector Laboratories Ltd. Peterborough, UK) were added onto a cover slip. The slide was placed onto the cover slip (chromosome side) and the DAPI solution was allowed to spread. After removing excess solution and ensuring no air bubbles were present the slide was sealed with nail varnish. Slides can be stored in the dark at -20°C

2.6.1.11 Image capture

Images were captured using a Leica epifluorescence microscope (Leica Microsystems Ltd. Milton Keynes, UK) equipped with cooled CCD camera (Photometrics, Tuscon, AZ, USA) and Smart Capture software (Digital Scientific Ltd. Cambridge, UK).

CHAPTER 3

Generation Of Mammalian Expression Vectors Containing Chromatin Modifying Elements

3.1 Introduction

Stable cell line generation using random integration is an unpredictable process, with only 1 in every 10,000 integration events being successful. Therefore, the probability of isolating a high expressing stable clone can be low and very labour intensive (Gorman and Bullock, 2000). The site of integration has a major effect on the transcription of a transgene. Integration into transcriptionally active regions of chromatin is preferential, however the majority of mammalian genomes are made up of heterochromatin and therefore the probability of the transgene being stably integrated into an area which is favourable for high and stable expression is low (Grewel and Elgin, 2002). Incorporation of chromatin modifying or opening elements into the transfected vector DNA attempts to circumvent this problem. The aim is to produce a stable environment for the integrant and subsequently to increase the probability of high level expression over a prolonged period of time.

The aim of the work described in this Thesis was to assess the ability of the chromatin modifying elements, UCOE, MAR, STAR and cHS4, to increase the level and stability of antibody expression in stable cell lines in the expression systems used at UCB. This Chapter will describe the construction of the vectors used.

In order to analyse the impact of the presence of various chromatin modifying elements on recombinant antibody expression a series of constructs was assembled. The elements were cloned into various locations within the vector surrounding the immunoglobin heavy chain (HC) and light chain (LC).

In all 28 vectors were made; 12 vectors were made with individual elements in 3 different locations, either upstream of the LC (5' LC) downstream of the HC (3' HC) or between the LC and HC (5' HC) (see Figure 3.7). A further 12 vectors were made which contained various combinations of 2 identical elements (5'LC 5' HC, 5' HC 3' HC, 3'HC 5' LC) (see Figures 3.8 and 3.9) and finally 4 vectors were constructed which contained the individual elements in all 3 locations (3' HC, 5' LC 5' HC) (see Figure 3.10).

To facilitate cloning a strategy was devised which made use of the unique *Mlu*I, *Not*I and *Sal*I sites within the vector. Incorporation of pairs of restriction enzyme sites with compatible cohesive ends at the 5' and 3' ends of the elements allowed the recreation of the unique restriction sites when the various fragments were cloned at the 3 different locations (see Figure 3.6).

3.1.1 Description of elements used in study

In this study it was decided to evaluate the smallest UCOE from the *HNRPA2B1-CBX3* locus, the core 1.5kb element (1.5kb A2UCOE; Antoniou *et al.*, 2003; Williams *et al.*, 2005). This element has been shown to be sufficient to produce the desired chromatin opening effect (Williams *et al.*, 2005) and its small size would be advantageous in vector manipulation. The 1.5kb A2UCOE contains the dual divergently linked promoters and approximately half of the CGI as shown in Figure 3.1.

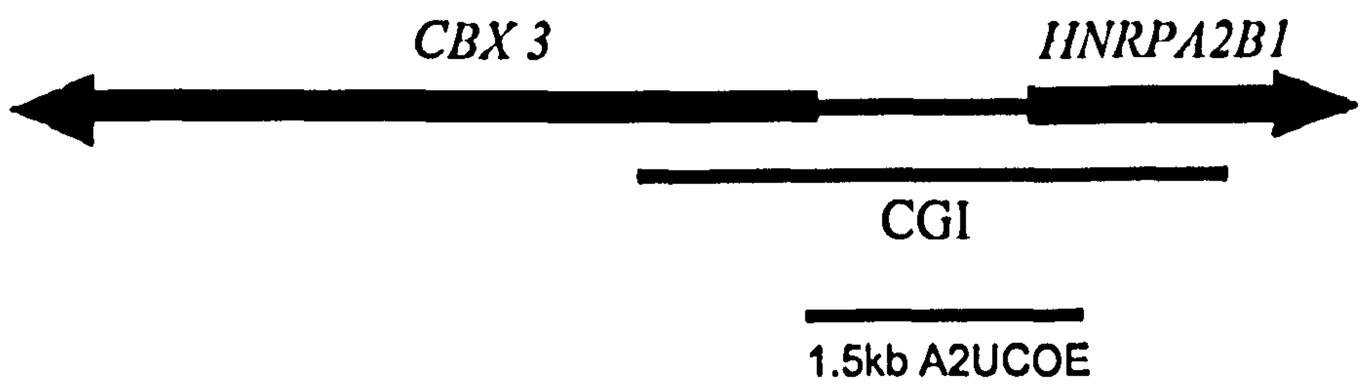


Figure 3.1 The human *IINRPA2B1/CBX3* locus and location of the 1.5kb UCOE used. Organisation of the *RNP* genomic locus indicating the locations of the divergent promoters (arrows), and the location of the 3 kb unmethylated CGI (black and white line). The fragment of the locus defined as the 1.5kb A2UCOE is indicated by the solid black line (adapted from Williams *et al.* 2005).

The sequence was analysed using the EMBOSS CpGplot software to determine the location and size of the CGI within the 1.5kb UCOE and the results are illustrated in Figure 3.2. The average percentage of C and G residues within the sequence is 64% and Figure 3.2C indicates that there are two CpG island regions within this sequence.

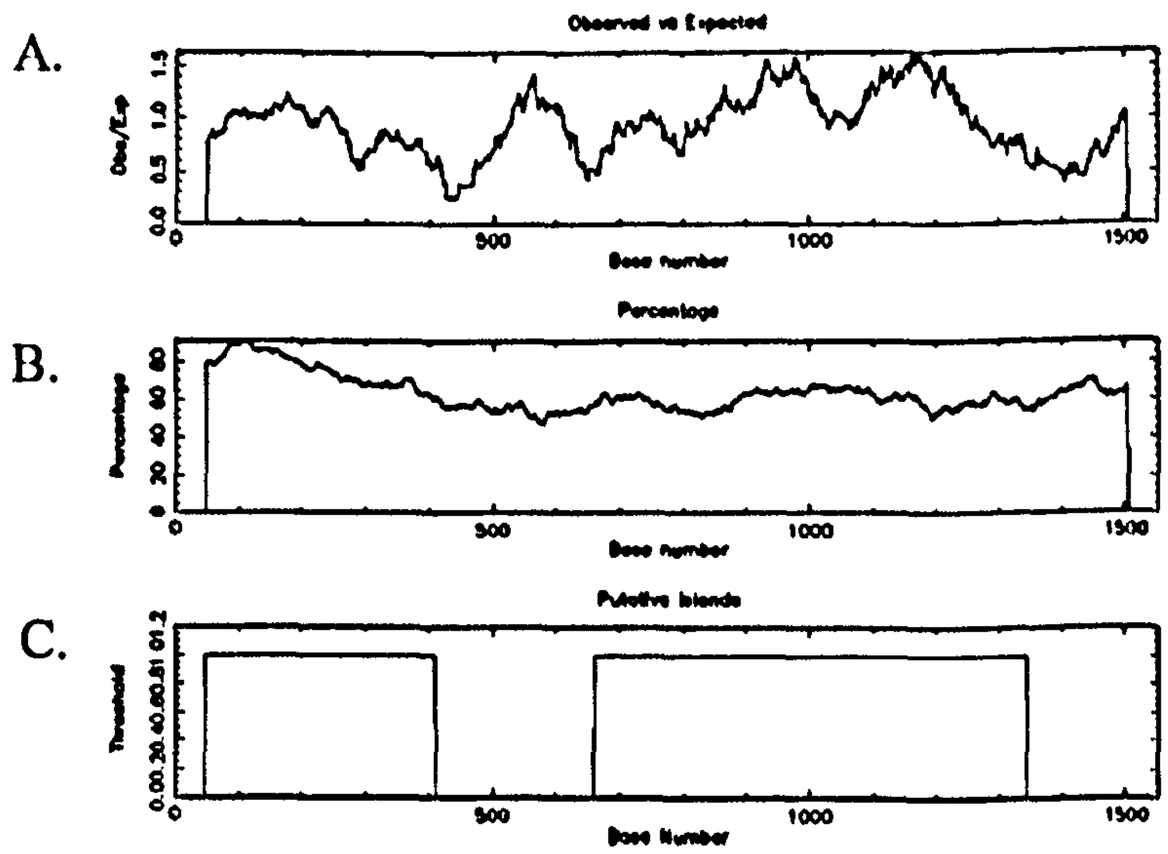


Figure 3.2 1.5kb UCOE Sequence Results From EMBOSS CpG Plot
A. Results from the observed to expected ratio of C plus G to CpG within the sequence. B. The percentage of G+C residues in the sequence window (sequence region which is moved along the sequence) and C. The putative CpG island regions in the sequence where cpgplot defines a CpG island as a region where, over an average of 10 windows and not less than 200 bases, the calculated (%G + %C) content is over 50% and the calculated Observed/Expected ratio is over 0.6

The MAR element investigated in the work described in this thesis was MAR X_S29. This MAR was identified in a bioinformatic search of the human genome (Girod *et al.* 2005). Data presented in the resultant Selexis patent application (WO2005/040377) demonstrated that incorporation of the identified MAR elements into expression vectors resulted in an increase in GFP expression from transfected CHO cells and that MAR X_S29 gave the highest increase. BLAST analysis indicated that MAR X_S29 is located on chromosome X. The BLAST analysis also indicated that this sequence is part of an intron in the *leucine rich repeats and calponin homology domain containing 2* (LRCH2) gene. Analysis of the sequence with the MAR-Wiz software (http://genomecluster.secs.oakland.edu/marwiz, Singh *et al.* 1997) which identifies sequences with characteristic MAR features such as origins of replication, TG-rich sequences, curved DNA, kinked DNA, topoisomerase II sites, and AT-rich sequences, reveals that there is a region within the 3462bp fragment that shows a high degree of MAR potential (see Figure 3.3).

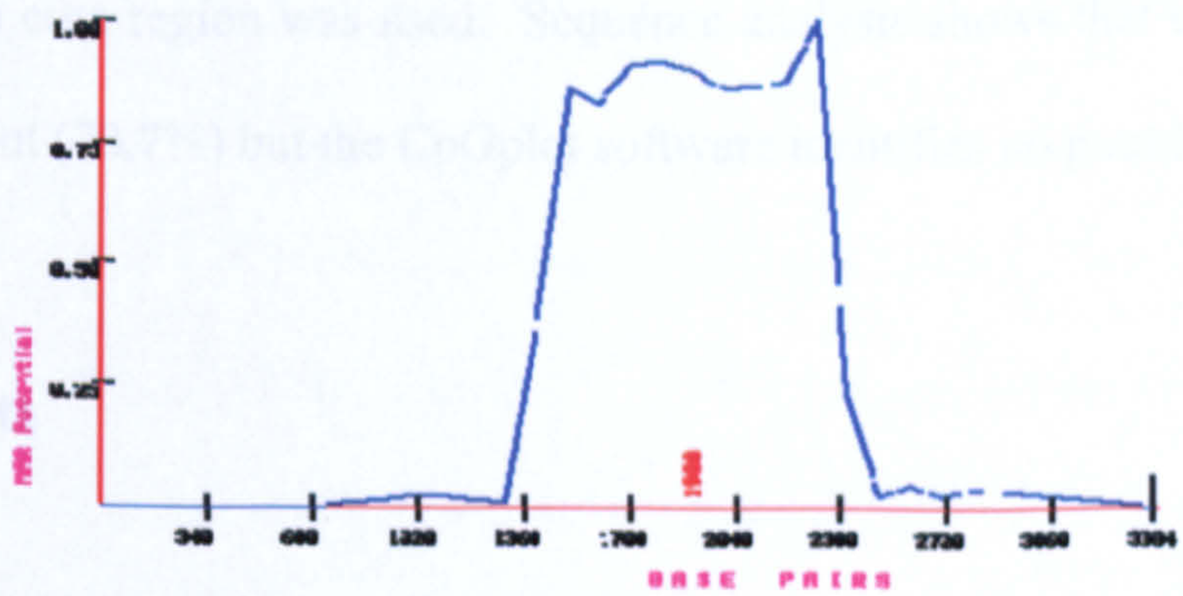


Figure 3.3 MAR X_S29 Sequence Analysed By MAR-wiz Software
DNA sequence for MAR X_S29 analysed using MAR-wiz software (http://genomecluster.secs.oakland.edu/marwiz) which illustrates the region of DNA which has a high degree of MAR potential.

The STAR elements chosen for investigation were STAR 40 and STAR 7. These elements, described by Kwaks *et al.* (2003) were identified from the human genome using a genetic screen. The cytogenic location of STAR 40 is chromosome 22, at q11.1 and it is located upstream of the *IL17R* gene. The result of BLAST searches concluded that the sequence was unique in the genome and that it encompassed entirely non-coding DNA. STAR 7 is located on chromosome 1 at q34 and is located in an area devoid of genes. The elements contain no distinct DNA motifs. Analysis of the STAR 40 and STAR 7 sequences using the CpGiplot software showed that there were no putative CpG islands in the sequences and no potential MAR regions were identified using the MAR-Wiz software.

The fragment of the chicken β-globin insulator chosen for this investigation was the 250bp core region. However, studies have shown that a transgene must be flanked by two copies of this core region in order to be effectively shielded from chromosomal effects (Pikaart et al., 1998) and therefore a tandem repeat (hereon referred to as HS4 tandem) of the core region was used. Sequence analysis shows that this region has a high GC content (70.7%) but the CpGiplot software identifies no putative CpG islands.

3.2 Results

3.2.1 Derivation of elements used in study

3.2.1.1 1.5kb A2UCOE

The 1.5kb A2UCOE was generated synthetically by overlapping oligonucleotide assembly and PCR by Entelechon. The fragment was designed to incorporate *Mlu*I and *Not*I restriction enzyme recognition sites at the 5' end and *Asc*I and *Eag*I

recognition sites at the 3' end to facilitate subsequent vector generation. At the same time, endogenous restriction sites for *Mlu*I, *Xho*I and *Eco*RI were removed, again, to facilitate vector generation. The sequence of the final 1.5kb A2UCOE can be seen in Appendix 1. A second version of the 1.5kb A2UCOE was generated by PCR to incorporate a 5' *Xho*I restriction enzyme site and a 3' *Sal*I site using the oligonucleotides listed in Appendix 2 (oligonucleotides 1.2.1 and 1.2.2). After successful amplification the PCR product was cloned into a TOPO vector to allow sequence verification using oligonucleotides 2.2.1 to 2.2.6, as shown in Appendix 2.

3.2.1.2 MARX_S29

The MARX_S29 sequence was obtained from the Selexis patent application (WO 2005/040377) and attempts were made to synthesise the element by overlapping oligonucleotide assembly. However, this proved to be unsuccessful presumably due to the high A/T content (the MAR element is 70.2% A/T) and the repetitive nature of the sequence. It was therefore decided to generate the MAR element from human gDNA by PCR. A BLAST search identified a PAC clone, RP4-736G20 which contained the entire sequence of MAR X_S29. Initial attempts to amplify the MAR element from the PAC clone template using oligonucleotides 1.1.1 and 1.1.2 (sequences in Appendix 2) under standard conditions (as described in Section 2.2.1) proved unsuccessful. Despite the fact that a variety of different combinations of amplifying oligonucleotides were used as shown in Figure 3.4A (oligonucleotides MAR F, MAR R and 1F to 6F, for sequences see Section 1.1 and 2.1, Appendix 2) no PCR products of the correct size were visible following agarose gel electrophoresis (Figure 3.4B). Similarly, reducing the annealing temperature to 50°C was equally unsuccessful. Standard PCR conditions use an extension temperature of 72°C,

however due to the A/T rich nature of the MAR sequence the temperature of the extension reaction was reduced to 60°C (Su *et al.*, 1996). This proved successful and a PCR fragment of the anticipated size was isolated (Figure 3.4C), digested with *Not*I and *Sal*I and cloned into pCR-Script for sequence confirmation. Two individual clones were sequenced, both of which had identical sequences, which corresponded to the sequence of MAR X_S29 provided in the patent application.

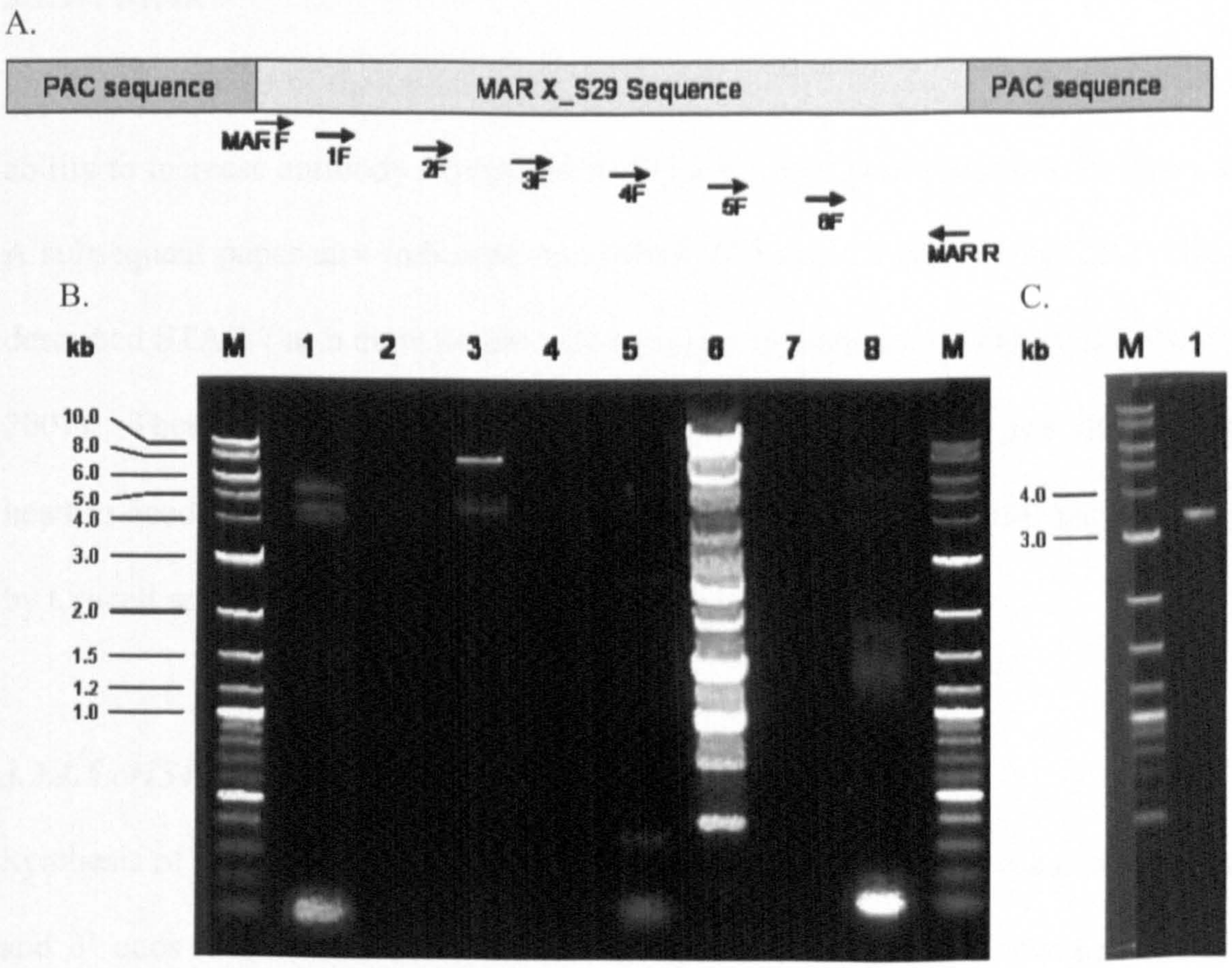


Figure 3.4 Amplification of MAR X_S29 from PAC Clone RP4-736G20 using internal oligonucleotides.

A. Location of oligonucleotides within MAR X_S29 sequence. B. Amplified PCR products analysed using agarose gel electrophoresis. Each forward oligonucleotide was paired with the MAR R oligonucleotide. M = DNA ladder, Lane 1: External oligonucleotides, Lane 2: 1F, Lane 3: 2F, Lane 4: 3F, Lane 5: 4F, Lane 6: 5F, Lane 7: 6F, Lane 8: External oligonucleotides – increased template (100ng). C. Successful amplification of MAR X_S29 using extension temperature of 60°C as analysed by agarose gel electrophoresis.

3.2.1.3 STAR 40

The sequence of the STAR 40 element was obtained from the publication by Kwaks *et al* 2003. The element was generated synthetically incorporating the required restriction enzyme sites at the 5' and 3' ends by overlapping oligonucleotide assembly and PCR by Entelechon.

3.2.1.4 STAR 7

During the course of the initial expression study STAR 40 failed to demonstrate any ability to increase antibody expression in CHO-K1 cells (see Section 4.2.1 and 4.2.2). A subsequent paper also indicated that STAR 40 has poor activity in CHO cells and described STAR 7 as a more potent anti-repressor element in this cell type (Otte *et al.*, 2007). Therefore, a vector with this element was generated for use in subsequent head-to-head transfections. This element was derived from a plasmid kindly supplied by Crucell and the sequence is shown in Appendix 1.

3.2.1.5 cHS4 tandem

Synthesis of the element incorporating the necessary restriction enzyme sites at the 5' and 3' ends was carried out synthetically by overlapping oligonucleotide assembly and PCR by Entelechon. The sequence of the tandem repeat used can be seen in Appendix 2.

3.2.1 Cloning strategy

The starting point of the antibody vector constructions was a double gene vector containing the antibody 535 genes (as described in Section 2.1.5 and shown in Figure 3.5). This vector contains unique restriction sites for: *Sal*I, downstream of the heavy

chain (3' HC); *Mlu*I, upstream of the light chain (5' LC) and *Not*I, located between the light and heavy chain (5' HC). The recognition sites for these enzymes have the same compatible cohesive end as a variety of other restriction enzymes. *Not*I has the same compatible overhang as *Eag*I, *Mlu*I and *Asc*I again share the same cohesive ends as do *Xho*I and *Sal*I. The individual elements were synthesised or amplified with these recognition sites at the 5' and 3' ends, as shown in Figure 3.6A. This allows the vector to be digested with the unique enzymes and the elements with both appropriate enzyme pairs. As depicted in Figure 3.6B, after the element has been cloned into the vector the recognition sequence for the unique *Mlu*I, *Not*I and *Sal*I sites are recreated and the correct orientation of the element can be distinguished by performing appropriate restriction digests.

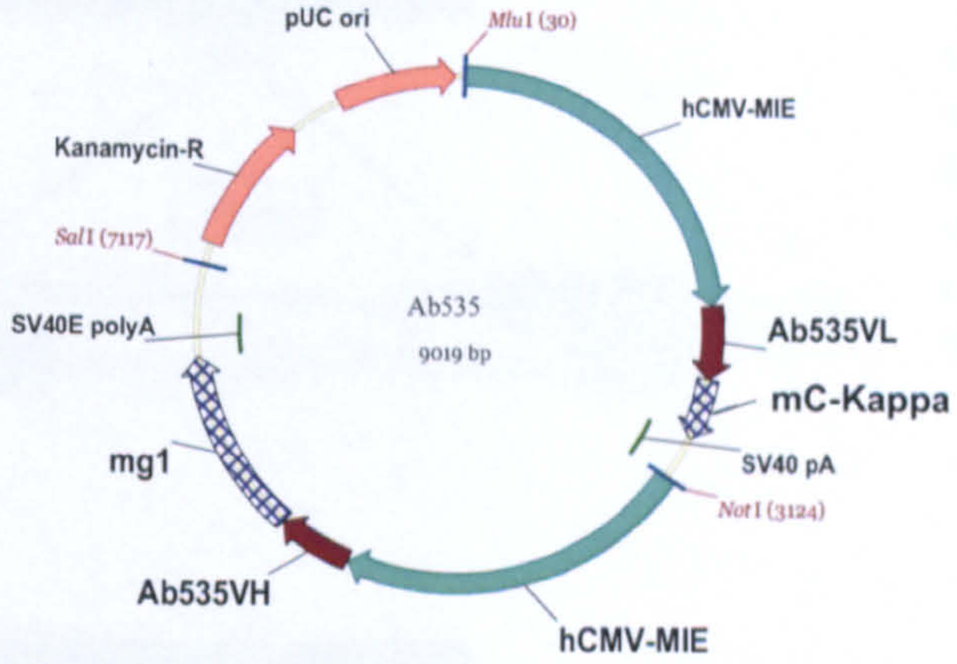
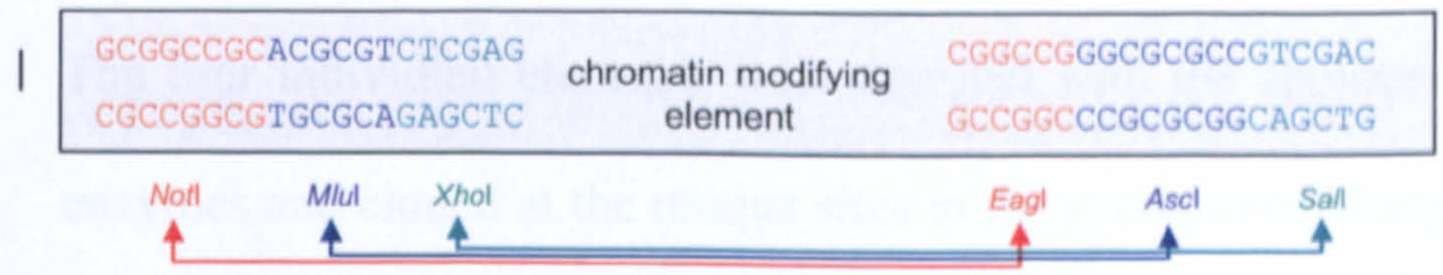
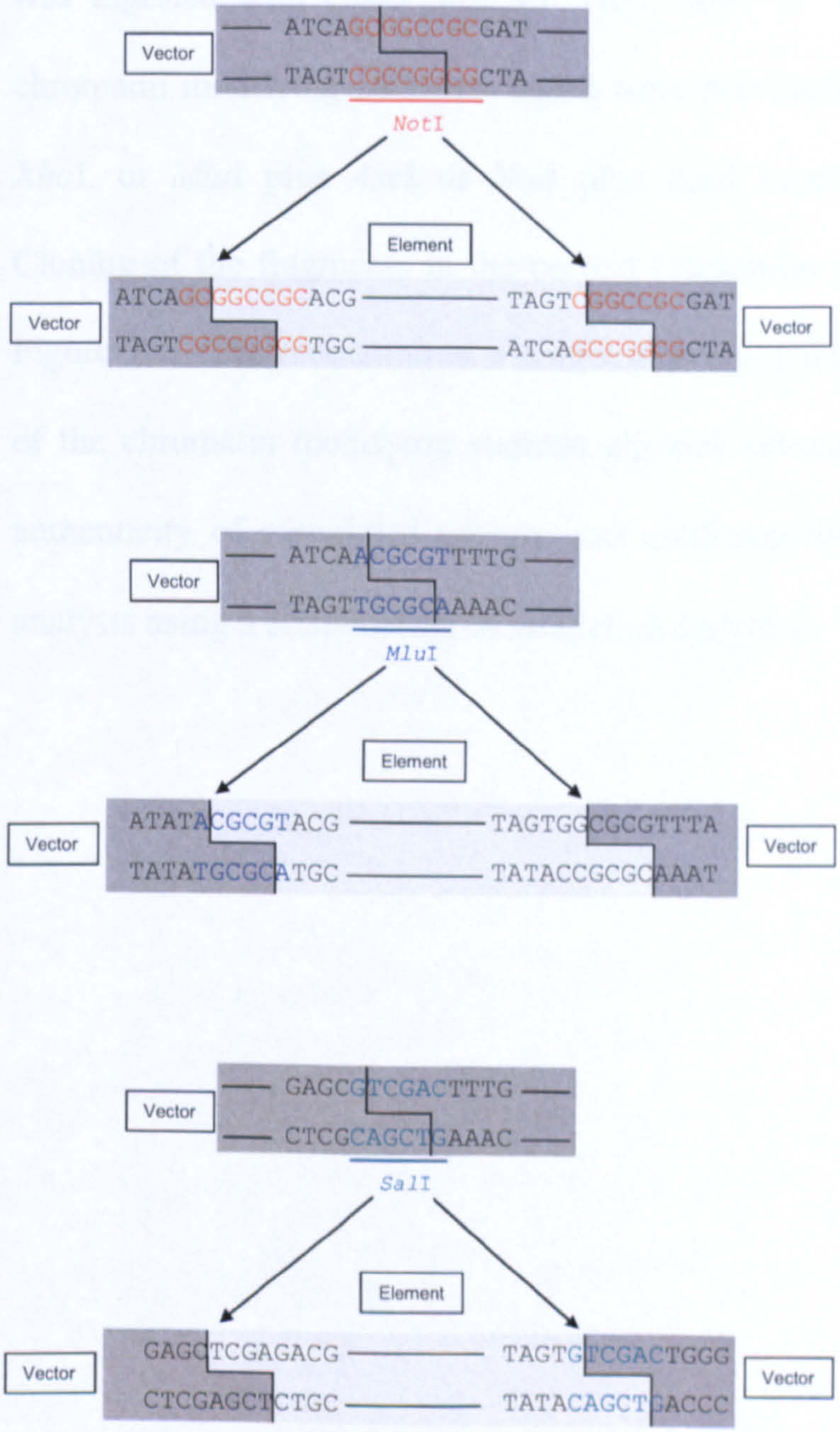


Figure 3.5 Ab535
In house mammalian expression vector containing Ab535 variable regions, murine Kappa and murine γ1 full length constant regions.

A.



B.



***Not*I site:**
The vector was digested with *Not*I and the element with *Not*I and *Eag*I. After ligation of the element into the vector the *Not*I site was recreated at the 5' end and the *Eag*I site remains at the 3' end.

***Mlu*I site:**
The vector was digested with *Mlu*I and the element with *Mlu*I and *Asc*I. After ligation of the element into the vector the *Mlu*I site was recreated at the 5' end and the *Asc*I site is lost.

***Sal*I site:**
The vector was digested with *Sal*I and the element with *Xho*I and *Sal*I. After ligation of the element into the vector the *Sal*I site was recreated at the 3' end and the *Xho*I site is lost.

Figure 3.6 Cloning strategy for vector generation.
A. Sequence of the 5' and 3' end of the chromatin modifying elements illustrating the position and sequence of the 5' and 3' restriction enzyme sites. B. Illustration of how unique *Not*I, *Mlu*I and *Sal*I sites are recreated.

3.2.3 Cloning of chromatin modifying elements into Ab535 vector

3.2.3.1 Single element vectors

The four individual elements were digested with the appropriate pairs of restriction enzymes and cloned at the unique sites in the vector to produce 12 vectors (3 for each element) with a single chromatin modifying element at each site. The Ab535 vector was digested with either *Sal*1 (3' HC), *Mlu*1 (5' LC), or *Not*1 (5' HC) and the chromatin modifying elements which were previously digested with either *Sal*I plus *Xho*I, or *Mlu*I plus *Asc*I or *Not*I plus *Eag*I combinations were ligated together. Cloning of the fragments in the correct orientation resulted in the vectors shown in Figure 3.7. The recreation of a unique restriction enzyme site at the appropriate end of the chromatin modifying element allowed subsequent vector manipulation. The authenticity of completed vectors was confirmed by diagnostic restriction enzyme analysis using a combination of restriction enzymes.

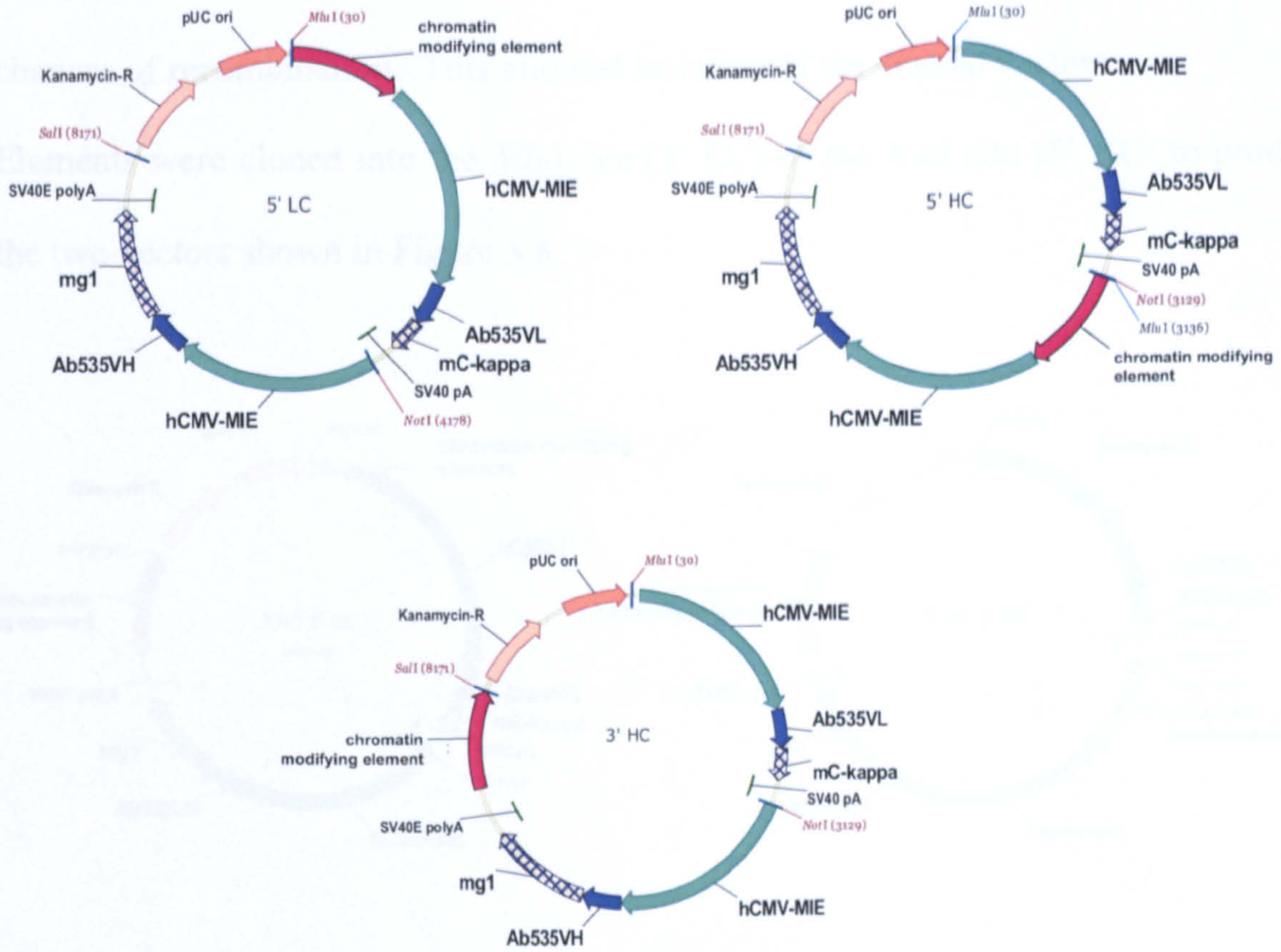


Figure 3.7 Vectors containing one copy of the chromatin modifying element.
Initial vectors were made with one chromatin remodelling element, either 3' to the heavy chain (3' HC), 5' to the light chain (5' LC) or 5' to the heavy chain (5'HC)

3.2.3.2 Double element vectors

The next stage was to clone a second element into vectors containing the element 3' to the heavy chain to make 8 vectors containing 2 identical elements. During this stage difficulties were encountered with the cloning of the elements and the desired constructs could not be isolated. It was presumed that the highly repetitive nature of the chromatin modifying elements and the fact that multiple copies of the elements were being incorporated into the same vector meant that internal vector recombination occurred during transformation of the XL1 blue supercompetent cells. An alternative *E. coli* strain, MAX Efficiency Stbl2™ cells, which had been designed for cloning unstable DNA sequences was subsequently used. In addition, the transformed

competent cells were propagated at 30°C instead of 37°C to further reduce the chances of recombination. This allowed isolation of the desired vectors.

Elements were cloned into the *Mlu*I site (5' LC) or the *Not*I site (5' HC) to produce the two vectors shown in Figure 3.8.

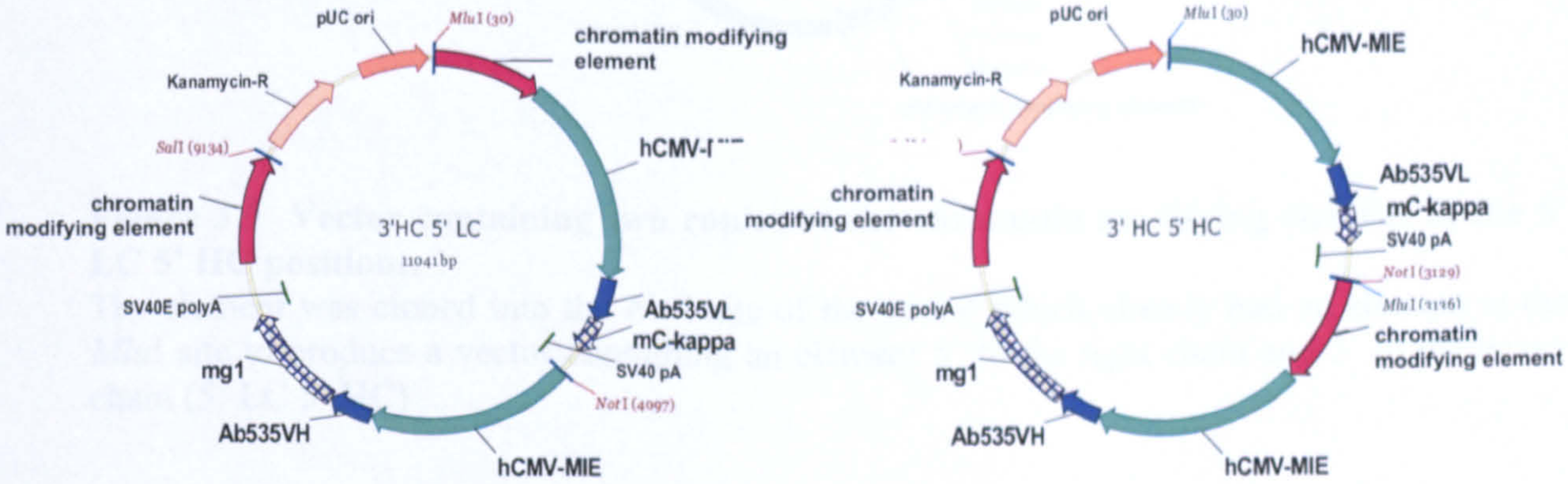


Figure 3.8 Vectors containing two copies of the chromatin modifying element
Elements were cloned into the *Mlu*I or *Not*I site of the 3' HC vector to produce vectors with elements 3' to the heavy chain and 5' to the light chain (3' HC 5' LC) and 3' to the heavy chain and 5' to the heavy chain (3' HC 5' HC)

Four further vectors containing two elements were made by cloning the second element into the *Not*I site of the 5' LC vector to create a vector containing an element 5' to both the HC and LC units (5' LC 5' HC) as illustrated in Figure 3.9.

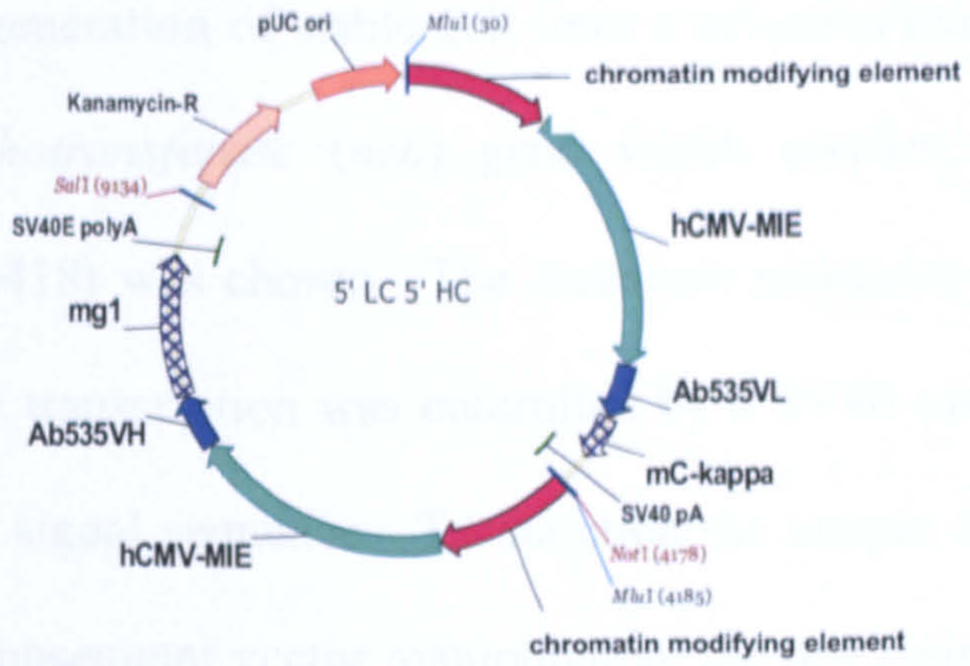


Figure 3.9 Vector containing two copies of the chromatin modifying element at the 5' LC 5' HC positions.
The element was cloned into the *Not*I site of the vector which already had an element at the *Mlu*I site to produce a vector containing an element 5' to the light chain and 5' to the heavy chain (5' LC 5' HC)

3.2.3.3 Triple element vectors

The final 4 vectors contained three copies of the chromatin modifying element at each unique site (*Mlu*I, *Not*I and *Sal*I). Elements were cloned into the *Not*I site of the 3' HC 5' LC vectors to produce vectors with elements 3' to the HC and 5' to both the HC and LC genes (3' HC 5' LC 5' HC) (Figure 3.10).

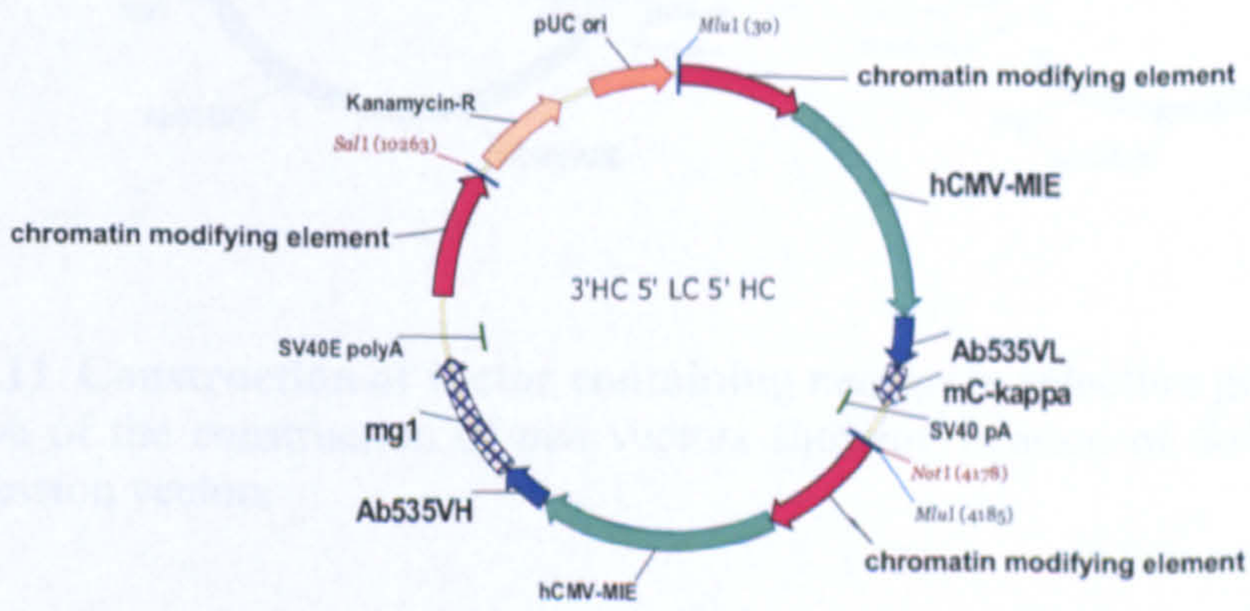


Figure 3.10 Vector containing three copies of the chromatin modifying element.
The element was cloned into the *Not*I site of the 3' HC 5' LC vector to produce a vector with three copies of the chromatin remodelling element at all three locations, 3' to the heavy chain, 5' to the light chain and 5' to the heavy chain (3'HC 5' LC 5' HC)

3.2.4 Cloning of Neomycin resistance marker

In order to allow the generation of stable cell lines a selection marker was required. The *neomycin phosphotransferase* (*neo*) gene which confers resistance to the antibiotic geneticin (G418) was chosen. The *neomycin* resistance cassette consisted of the *neo* gene whose transcription was controlled by a SV40 early promoter and a SV40 polyadenylation signal sequence. To maintain the unique *Sal*I site within the vectors and to aid in subsequent vector manipulation, the neo resistance cassette was amplified with a *Sal*I site at the 5' end and a *Xho*I site at the 3' end (see Appendix 2, oligonucleotides 1.3.1 and 1.3.2) as described in Section 2.2.1. After the sequence of the amplified neo resistance cassette was verified using oligonucleotides 2.3.1 to 2.3.2 (as shown in Appendix 2) the fragment was ligated into the vectors as depicted in Figure 3.11.

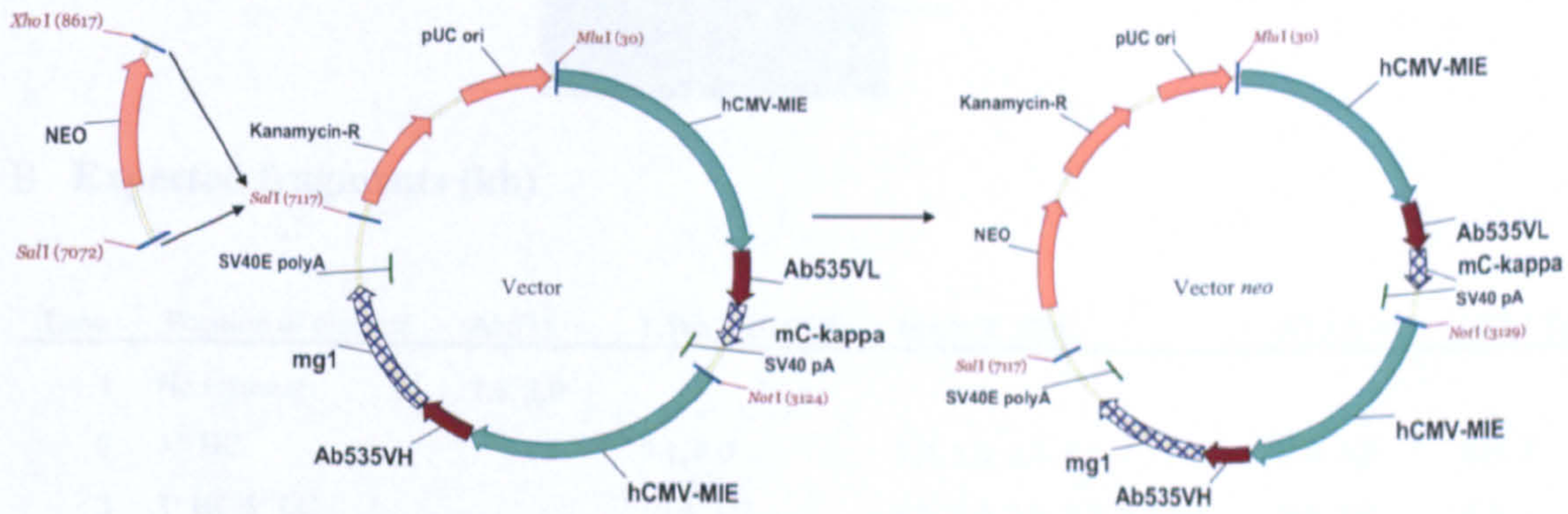
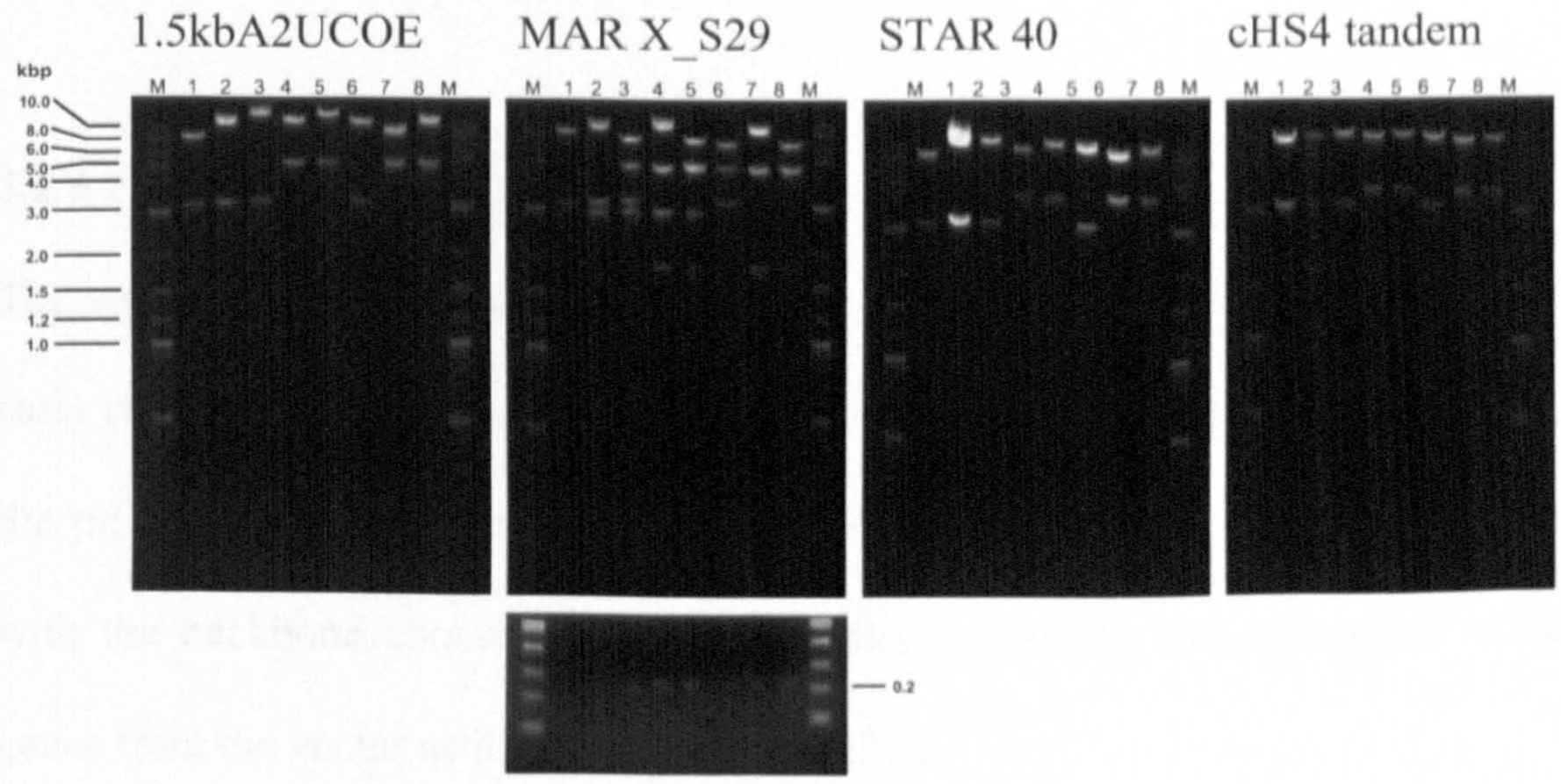


Figure 3.11 Construction of vector containing *neomycin* selection gene
Illustration of the construction of *neo* vectors showing ligation of *Sal*I-*Xho*I PCR fragment into expression vectors

3.2.5 Restriction digest analysis to confirm vector constructs

A total of 28 vectors were constructed, 7 for each element. The authenticity of the constructed vectors was confirmed by a set of diagnostic restriction enzyme digests. Digests were perfomed on all vectors to confirm that each construct was correct before stable cell line generation took place. An example of the results obtained following digestion of the 28 vectors with *Hin*dIII can be see in Figure 3.12.

A. **Agarose gel electrophoresis separation of digested DNA**



B. **Expected fragments (kb)**

Lane	Position of element	Ab535	1.5kb A2UCOE	MAR X_S29	STAR 40	cHS4 Tandem
1	No element	7.5, 3.0				
2	3' HC		9.1, 3.0	7.9, 3.0, 2.7, 0.2	8.4, 3.0	8.0, 3
3	3' HC 5' LC		10.6, 3.0	6.6, 4.4, 3.0, 2.7, 0.2 x 2	9.6, 3.0	8.5, 3
4	3' HC 5' HC		9.1, 4.6	7.9, 4.4, 2.7, 1.8, 0.2 x 2	8.5, 4.1,	8.0, 3.5
5	3' HC 5' HC 5' LC		10.7, 4.6	6.6, 4.4, 4.3, 2.7, 1.8, 0.2 x 3	9.6, 4.1	8.4, 3.5
6	5' LC		9.1, 3.0	6.2, 4.4, 3.0, 0.2	8.6, 3.0	8.0, 3
7	5' HC		7.5, 4.6	7.5, 4.4, 1.8, 0.2	7.5, 4.1	7.5, 3.5
8	5' HC 5' LC		9.1, 4.6	6.2, 4.4, 4.4, 1.8, 0.2 x 2	8.5, 4.1	8.0, 3.5

Figure 3.12 Restriction digests confirming correct cloning of elements
500ng of DNA was digested with the restriction enzyme *Hin*dIII for 1.5 hours at 37°C. Reactions were analysed by gel electrophoresis. A. Fragments generated from digests for each vector. B. Approximate fragment sizes for expected correct vector constructs. (kb).

3.2.6 Changing selection system

The results of the initial expression studies to determine the optimal vector configurations demonstrated that high expressing clones could not be isolated using the *neomycin* selection gene. However, expression levels were sufficient to select the optimal configuration of the individual elements (see Sections 4.2.1 to 4.2.3). To progress the work further and isolate cell lines with increased expression levels the selection system was changed from neomycin to glutamine synthetase (GS) (licensed from Lonza) for subsequent studies.

3.2.6.1 Ab535

The vector backbone which consisted of the SV40 early promoter, GS cDNA, SV40 early region polyadenylation signal and the *ampicillin* resistance gene was taken from the pEE12 vector (Bebbington 1995) as a *Sal*I-*Mlu*I fragment. This was exchanged with the backbone containing the neo resistance cassette and *kanamycin* selection genes from the vector as illustrated in Figure 3.13.

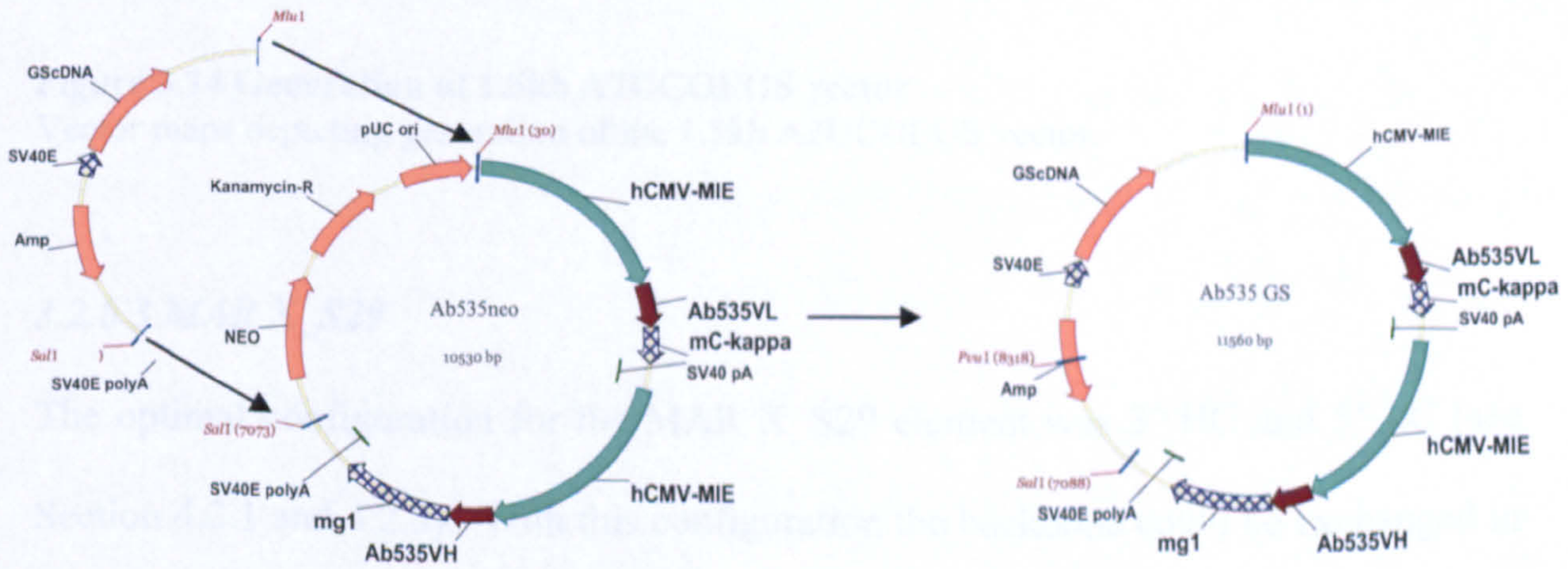


Figure 3.13 Construction of Ab535 vector containing glutamine synthetase selection marker

The *Sal*I-*Mlu*I region of the vector backbone from the Ab535neo vector was exchanged with the *Sal*I-*Mlu*I region from the pEE12 vector which contains the glutamine synthetase selection marker and *ampicillin* resistance gene.

3.2.6.2 1.5kb A2UCOE

The optimal vector for 1.5kb A2UCOE was 3' HC 5' LC' 5'HC (see Section 4.2.1 and 4.2.2). To exchange the vector backbone to incorporate the GS selection marker a three way ligation was conducted. This was due to the presence of a *Mlu*I site within the 1.5kb A2UCOE at the 3' HC site and the strategy used is depicted in Figure 3.14. The *Mlu*I-*Bsi*WI, *Bsi*WI-*Sal*I fragments from the 1.5kb A2UCOEneo vector were ligated together with the *Sal*I-*Mlu*I fragment from pEE12 to generate the 1.5kb A2UCOEGS vector.

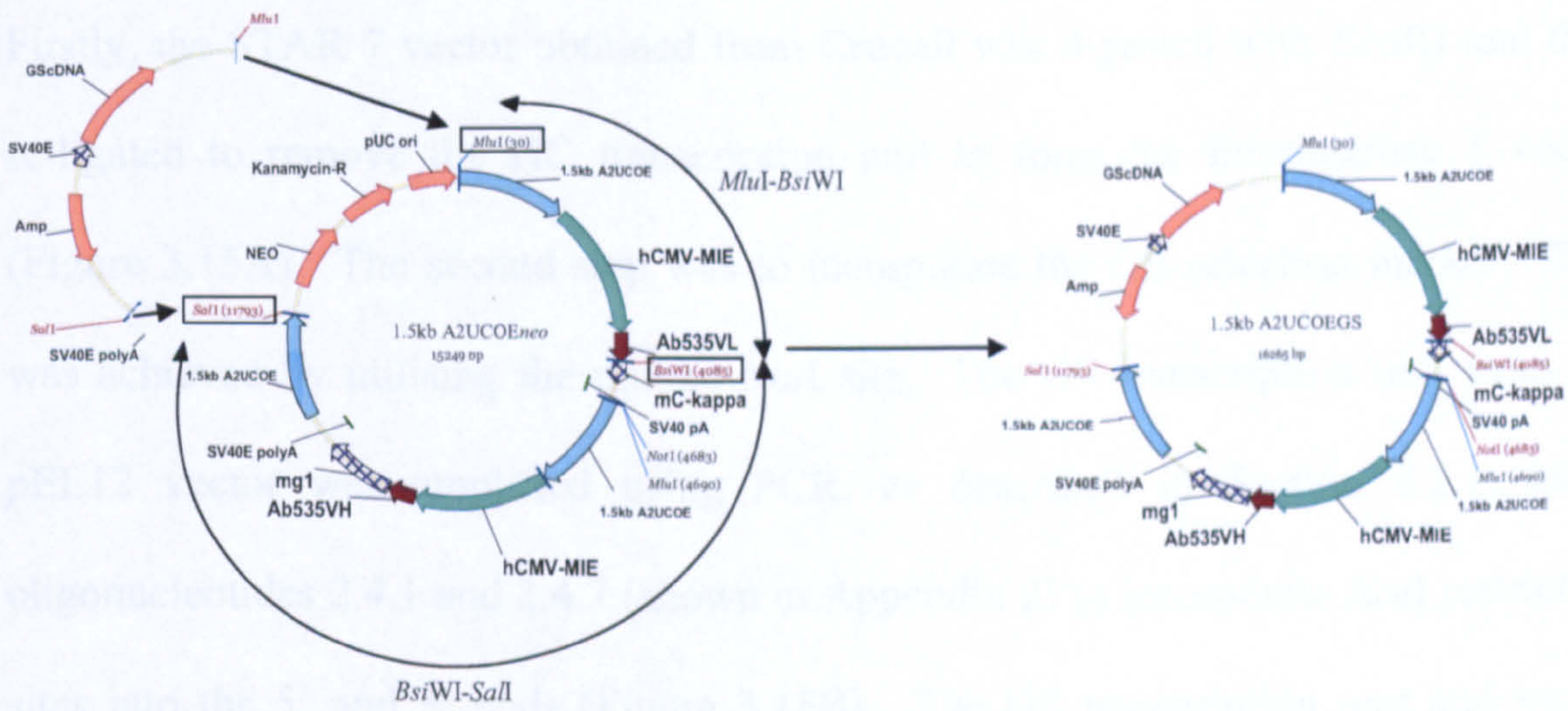


Figure 3.14 Generation of 1.5kb A2UCOEGS vector
Vector maps depicting generation of the 1.5kb A2UCOEGS vector.

3.2.6.3 MAR X_S29

The optimal configuration for the MAR X_S29 element was 3' HC and 5' LC (see Section 4.2.1 and 4.2.3). With this configuration the backbone could be exchanged in the same way as the Ab535 vector.

3.2.6.4 STAR 7

During the course of the initial expression study STAR 40 failed to demonstrate any ability to increase antibody expression in CHO-K1 cells (see Section 4.2.1 and 4.2.2). The subsequent paper by Otte *et al* also indicated that STAR 40 has poor activity in CHO cells. The paper described STAR 7 as a more potent anti-repressor element in this cell type (Otte *et al.*, 2007). Therefore a vector was constructed with the element at the locations described in the paper which was 3' to the HC and 5' to the LC.

The STAR 7 GS vector was made in three steps, which is depicted in Figure 3.13. Firstly, the STAR 7 vector obtained from Crucell was digested with *Eco*RI and then re-ligated to remove the HC transcription unit to form the Intermediate 1 vector (Figure 3.15A). The second step was to incorporate the GS selection marker. This was achieved by utilising the unique *Not*I site. The GS transcription unit from the pEE12 vector was amplified using PCR, as described in Section 2.2.1, using oligonucleotides 2.4.1 and 2.4.7 (shown in Appendix 2) to incorporate *Not*I restriction sites into the 5' and 3' ends (Figure 3.15B). The GS transcription unit and vector were both digested with *Not*I and the GS fragment was ligated into the Intermediate 1 vector generating the Intermediate 2 vector. The GS transcription unit was sequence verified using oligonucleotides 2.4.1 to 2.4.12 (as shown in Appendix 2) to confirm no mutations had been incorporated into the sequence during PCR amplification. The final step of the cloning was to incorporate the HC and LC transcription units from Ab535 into the vector (Figure 3.15C). This was achieved by utilising the *Mlu*I and *Bam*HI sites within both vectors. The Intermediate 2 vector was digested with *Mlu*I and *Bam*HI to generate a fragment containing the two STAR 7 elements, the *ampicillin* resistance gene and the GS transcription unit. The presence of a *Bam*HI

site within the murine γ1 constant region meant that three fragments (*Mlu*I-*Bam*HI, *Bam*HI-*Xma*I, *Xma*I-*Bam*HI, see Figure 3.15C) had to be generated utilising the *Xma*I site between the two *Bam*HI sites. A four-way ligation was carried out which generated the final STAR7GS vector (Figure 3.15D).

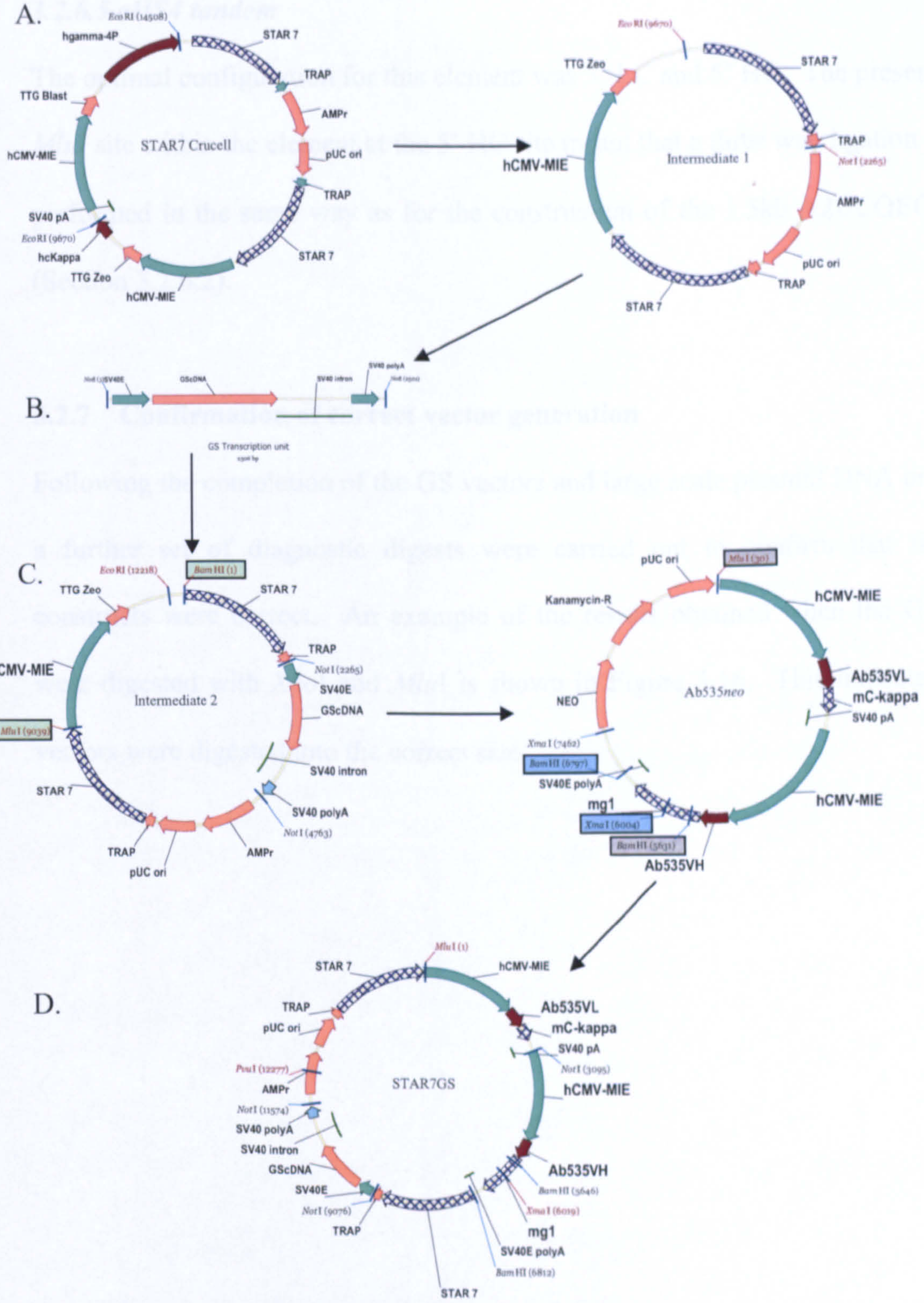


Figure 3.15 Cloning strategy to make STAR 7 GS vector

A. The STAR7 Crucell vector was digested with *Eco*RI to remove the heavy chain transcription unit. B. The PCR amplified GS transcription unit with *Not*I sites at 5' and 3' ends was ligated into Intermediate 1 vector to form Intermediate 2. C. Four way ligation to make final vector containing: *Bam*HI-*Mlu*I fragment from intermediate 2 and *Mlu*I-*Bam*HI, *Bam*HI-*Xma*I and *Xma*I-*Bam*HI fragments from Ab535neo vector. D. Final STAR7GS vector.

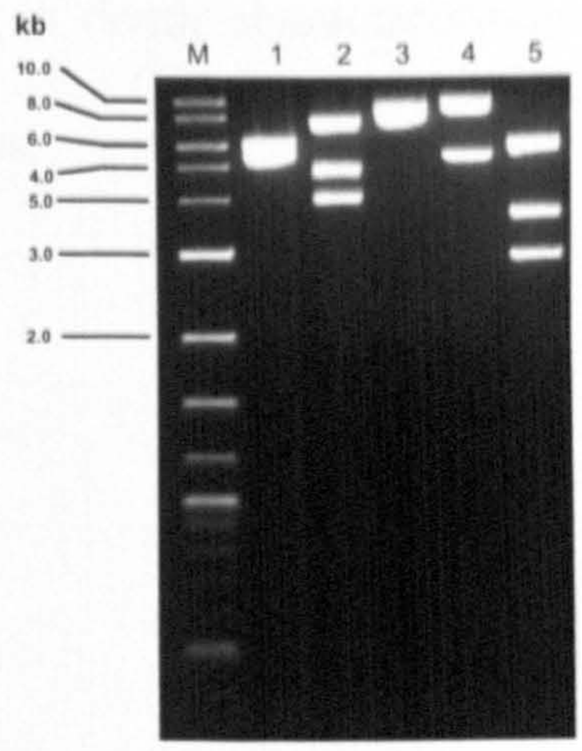
3.2.6.5 cHS4 tandem

The optimal configuration for this element was 5' LC and 5' HC. The presence of the *Mlu*I site within the element at the 5' HC site meant that a three way ligation had to be performed in the same way as for the construction of the 1.5kb A2UCOEGS vector (Section 3.2.6.2).

3.2.7 Confirmation of correct vector generation

Following the completion of the GS vectors and large scale plasmid DNA preparation a further set of diagnostic digests were carried out to confirm that the vector constructs were correct. An example of the results obtained when the GS vectors were digested with *Xho*I and *Mlu*I is shown in Figure 3.16. This indicated that all vectors were digested into the correct size.

A.



B.

	Construct	Expected fragments (kb)
1	Ab535	6.0, 5.5
2	1.5kb A2UCOE	7.6, 4.6, 3.9
3	MAR X_S29	9.4, 8.8
4	STAR 7	10.3, 5.5
5	cHS4 tandem	6.0, 3.5, 2.9

Figure 3.16 Restriction digests confirming correct cloning of GS vectors
500ng of DNA was digested with the restriction enzymes *Xho*I and *Mlu*I for 1.5 hours at 37°C. Reactions were analysed by gel electrophoresis. A. Fragments generated from digests. B. Expected fragment size if vector construct was correct.

3.3 Discussion

This Chapter has described the derivation of the chromatin modifying elements utilised in the study and the cloning strategy used to generate the panel of constructs containing these elements at various positions within the Ab535 double gene vector. The inclusion of an antibiotic selection marker within the vectors allowed for efficient generation of stable cell lines to assess the ability of the chromatin modifying elements to increase antibody expression in stable cell lines. Having established the optimal combination for the individual elements, as described in subsequent Chapters, further cloning was carried out to change the selection marker in order to generate

vectors required for the in depth characterisation of stable cell lines derived from these optimised constructs.

CHAPTER 4

Assessing The Ability Of Chromatin Modifying Elements To Increase Antibody Expression In CHO-K1 Stable Cell Lines

4.1 Introduction

A series of expression constructs were generated to assess the effect of the various chromatin modifying elements on antibody expression in stable cell lines (described in Chapter 3). Initially seven different vectors were generated for each element, which contained different numbers and different combinations of the individual elements in various positions in the vector. These constructs were used to stably transfect CHO–K1 cells and antibody expression levels were assessed either as pooled stable cell populations or individual clones using a neomycin selection system.

For the generation of pooled stables the whole transfected cell population is grown as a single culture, the advantage of this approach is that it generates recombinant protein quickly, compared to having to isolate clonal cell lines. In this method, following transfection, selective pressure is applied and as expected cell viability and cell number falls (an example of a typical growth curve is shown in Figure 4.7). After a period of time the transfected drug resistant cells then start to grow and cell viability and cell numbers begin to recover in the population. The pooled culture then consists of a mixed population of individual cell lines with a range of expression rates. The expression level within the pool represents an average of these individual cell lines. Over time the pooled stable cell line may become less diverse as individual cell lines with enhanced growth characteristics out-grow other cell lines within the population.

Having determined the optimal configuration for the individual elements a more detailed study was undertaken. For this subsequent work the selection system was changed from neomycin to glutamine synthetase. The effect of the chromatin

modifying elements on stable cell line generation was again studied as both pooled and clonal cell lines.

4.2 Results

4.2.1 Initial expression studies to determine optimal vector configurations: Analysis of pooled stable cell lines

In order to establish the optimal location and combination for individual chromatin modifying elements seven constructs for each of the different elements (positioned in the combinations described in Section 3.2.3), in addition to a control vector which contained no elements, were stably transfected into CHO-K1 cells as described in Section 2.4.3.1. For ease of analysis this investigation was performed as pooled stables.

At two weeks post transfection when cell populations had fully recovered to higher than 98% viability and cell numbers were sufficiently high, overgrown cultures were set up to assess antibody productivity. Cells were seeded at 3 x 10^5 cells/ml into 125ml Erlenmeyer flasks and grown in batch culture for 10 days at which point viability had decreased to 10-20%. Supernatant was removed and assayed by mouse IgG ELISA as described in Section 2.5.1 to determine expression levels and results are shown in Figure 4.1. The highest increase in expression over the Ab535 control construct was observed with constructs containing the 1.5kb A2UCOE (Figure 4.1A), with the presence of two elements (5' LC 5' HC) resulting in a 3-fold increase in expression, whereas the presence of three elements (3' HC 5' LC 5' HC,) gave a 6.5-fold increase over those obtained with the control Ab535 expression vector. The

remaining three elements (MAR X_S29, STAR 40 and cHS4 tandem did not increase expression levels to such an extent, with at best a 2-fold increase observed in the case of MAR X_S29 (Figure 4.1 B).

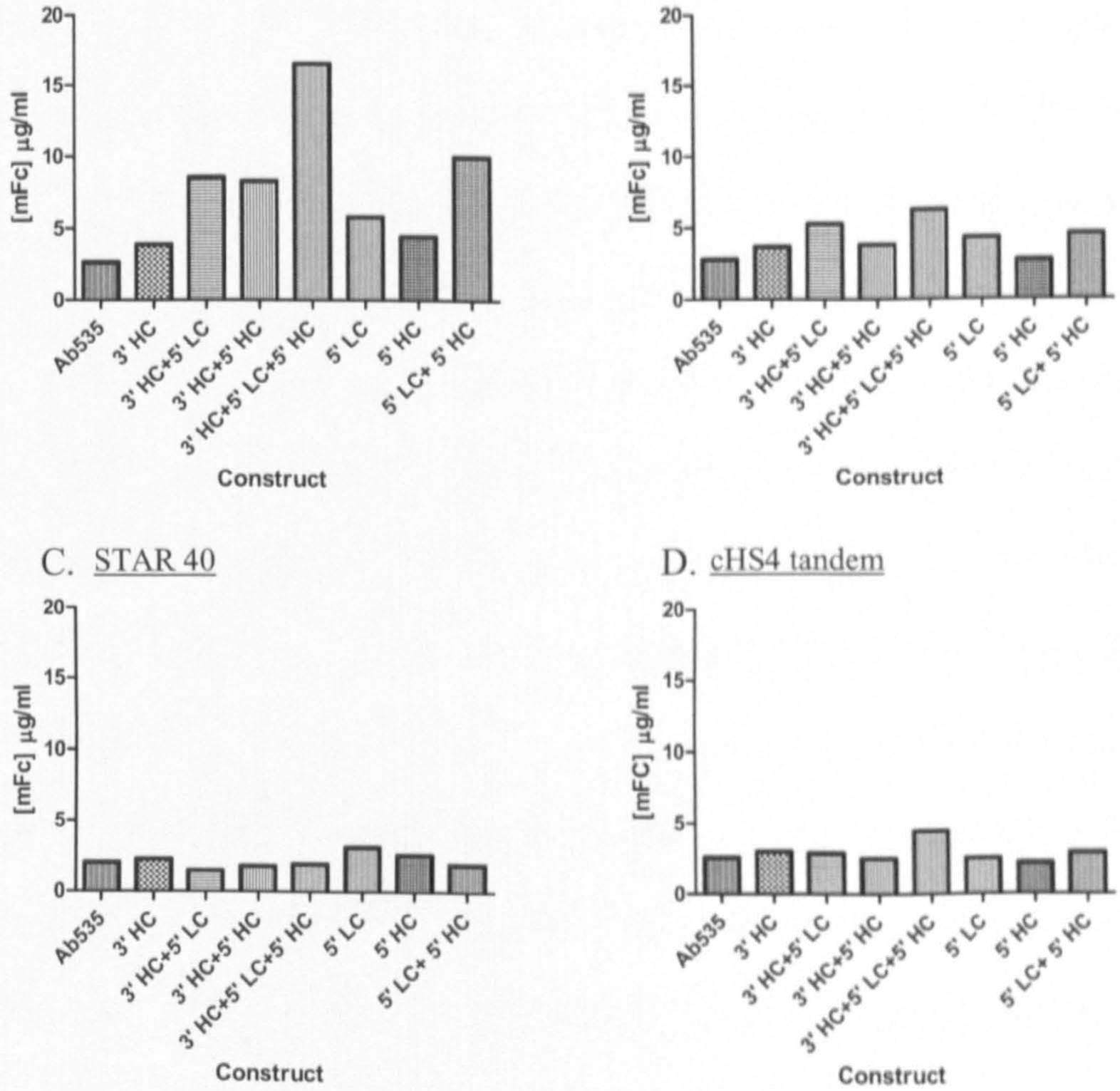


Figure 4.1 Expression levels of CHO-K1 pooled stables +/- chromatin modifying elements

Cells were seeded at $3x10^5$ cells/ml expression was measured after 10 days using a mouse IgG ELISA. MOPC21 was used as the standard at a starting concentration of 1µg/ml. Duplicate samples from each pooled stable were added to different ELISA plates at a 1 in 10 starting dilution and then titrated down the plate using tripling dilutions. Each data point is representative of the mean. Data shown is from one experiment.
A. 1.5kb UCOE, B. MAR X_S29, C. STAR 40, D. HS4 tandem.

4.2.2 Initial expression studies to determine optimal vector configurations: Analysis of clonal stable cell lines

In order to isolate clonal stable cell lines the pooled stable cell lines were plated out in CloneMedia, (see Section 2.4.4.1). Cells were grown for 10 days and subsequently

picked using an automated colony picker (Clonepix^{FL}) using white light. The isolation of single clones would allow the frequency of expressing clones within the population for each construct to be assessed. A total of 192 colonies were picked for each construct into two 96 well plates and cultured for 7 days. The relative expression and the distribution of expressing clones for each construct was ascertained using spot ELISA (as described in Section 2.5.2). In the first instance only the relative distribution of expressing cell lines was being analysed so only relative rather than absolute expression levels were required for each clone on the 96 well plate. Wells were visualised and scored for cell growth and absorbance values from the ELISA were only considered from wells positive for growth. An example of the data obtained from the spot ELISAs is shown in Figure 4.2. The distribution of expressing clones for each plate was calculated as described in Figure 4.3. Due to each plate having a different number of wells that scored positive for cell growth the numbers of clones per quartile were presented as a percentage of the total number of clones from the two plates. Data for each construct can be found in Appendix 3. Using this calculation the percentage of clones in the 1st quartile could be determined for each configuration of the chromatin modifying elements under study as shown in Figure 4.4.

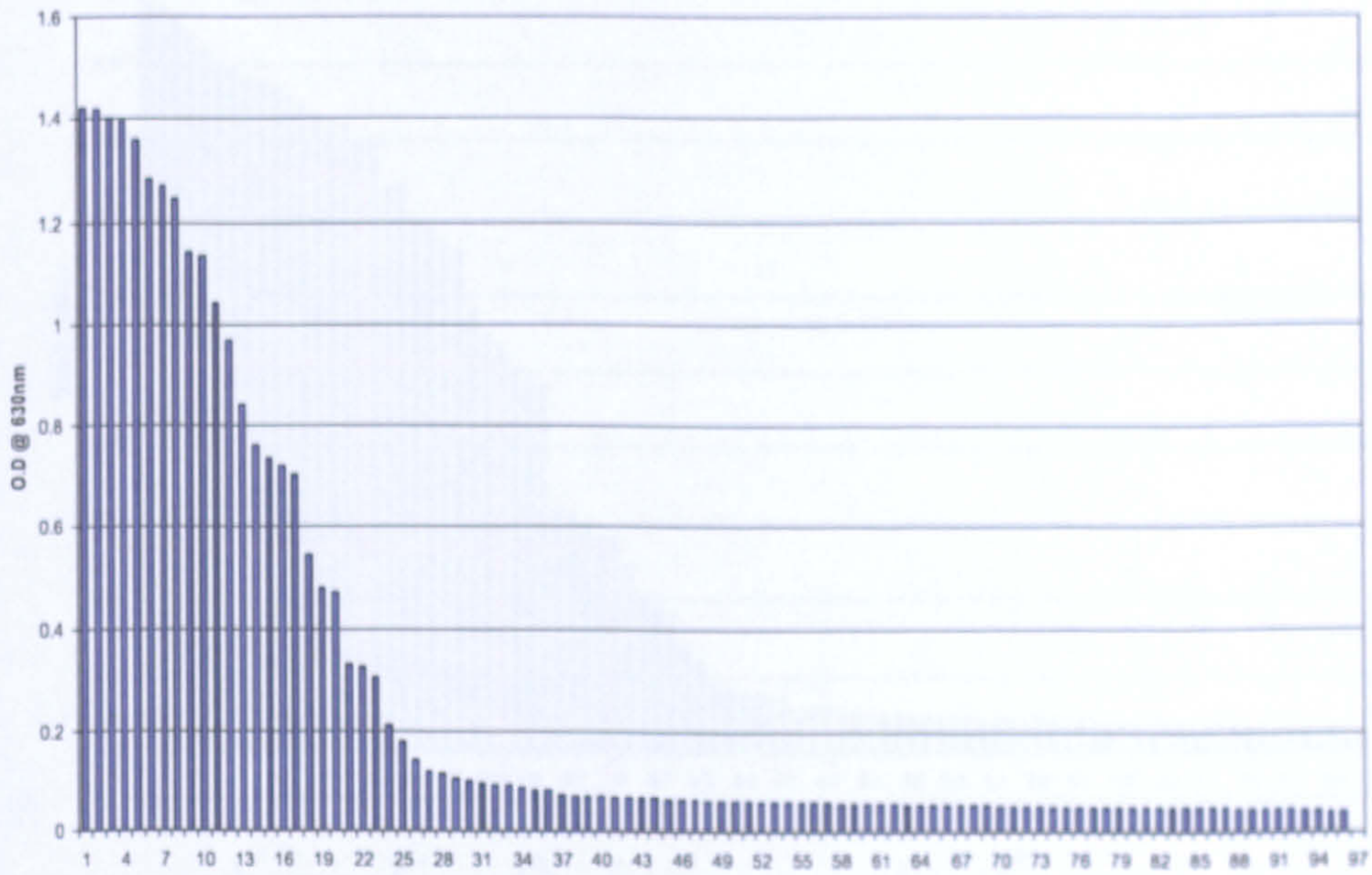
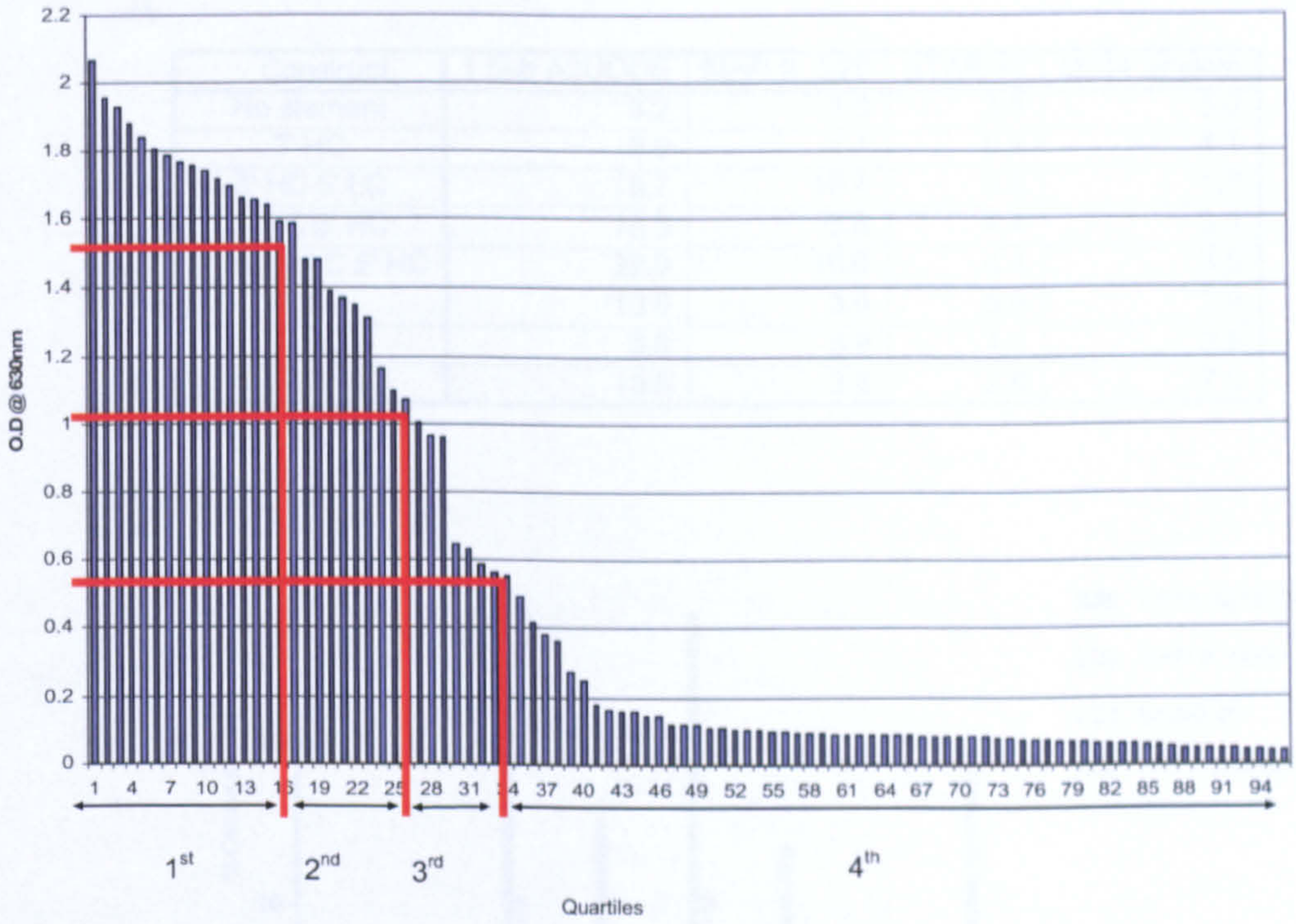


Figure 4.2 Representative data from spot ELISA
For each 96 well plate picked the absorbance values were ranked from highest to lowest. Data represents absorbance values from 1.5kb A2UCOE (5' HC 5' LC) clones

A



B: 2.066-0.042 = 2.024

C: 2.024÷4 = 0.506

D: 2.066 ——— 1.56 ——— 1.054 ——— 0.548 ——— 0.042

17 clones | 9 clones | 8 clones | 62 clones

Figure 4.3 Calculating the number of clones in each quartile

A. Graph illustrating the absorbance values for each clone from a 96 well plate with the quartile ranges depicted. These quartile ranges are calculated as follows.

The absorbance range is calculated by subtracting the lowest absorbance value (from a well positive for cell growth) from the upper value (Calculation B). This figure is then divided by 4 to give the size of each quartile (Calculation C). This value is subtracted from the highest absorbance value and then subsequently from the resulting values to give the cut off points for each quartile. The number of clones in each quartile is then determined (D).

Data represents absorbance values from 1.5kb A2UCOE (3'HC 5'LC) clones.

A

Construct	1.5kb A2UCOE	MAR X_S29	STAR 40	cHS4 tandem
No element	3.2	3.6	3.5	3.0
3' HC	3.9	4.1	6.4	4.1
3' HC 5' LC	18.7	10.4	5.6	1.7
3' HC 5' HC	15.3	3.8	4.4	3.3
3' HC 5' LC 5' HC	27.9	10.6	4.4	3.5
5' LC	13.8	3.9	5.5	2.9
5' HC	5.9	5.9	4.3	3.5
5' LC 5' HC	19.6	3.8	2.9	7.0

B

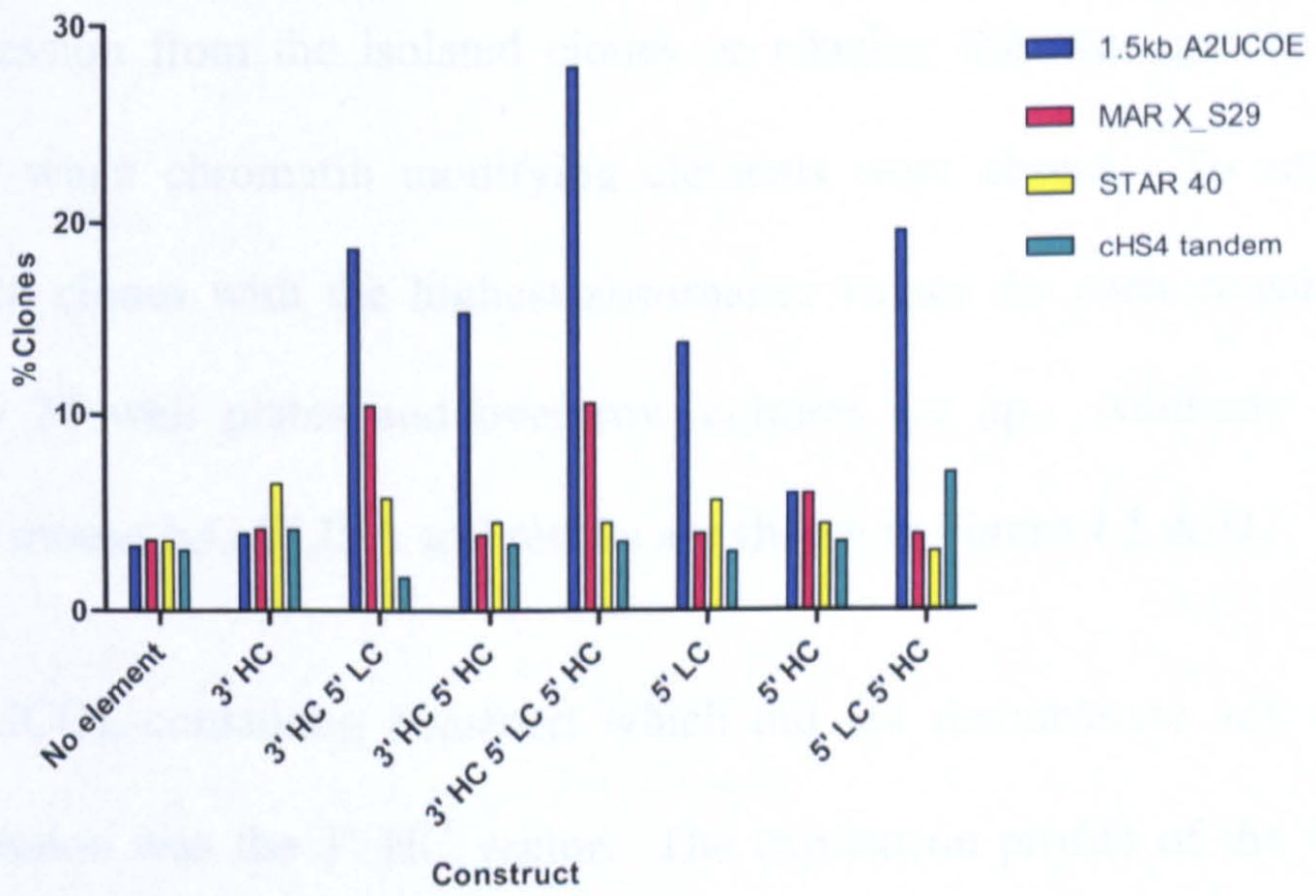


Figure 4.4 Frequency of highest expressing clones.
A. Table of results comparing the percentage of clones within the 1st quartile for each construct and element. B. Graph showing % of clones in 1st quartile for each element.

Figure 4.4 shows that the 1.5kb A2UCOE element gives the highest increase in the proportion of higher expressing clones compared to the other elements (Figure 4.4, blue bars). The largest increase was observed with the vector containing three 1.5kb A2UCOEs at the 3' HC 5' LC 5' HC positions. This resulted in a 8.6 fold increase in the number of clones in the first quartile. The MAR X_S29 element gave the second highest increase with two constructs (3' HC 5' LC and 3' HC 5' LC 5' HC) both showing an almost 3-fold increase over the control vector Ab535 (Figure 4.4, purple bars). The increase in the number of clones in the first quartile for the cHS4 tandem cell lines were less pronounced with the highest increase (2.3-fold) observed using the 5' LC 5' HC vector (Figure 4.4, green bars). The STAR 40 also only showed a

modest (<2 fold) increase in the number of highest expressing clones (Figure 4.4, yellow bars) which is consistent with the data from the individual pooled stable cell lines (see Section 4.2.1).

The presence of the 1.5kb A2UCOE and MAR X_S29 elements increased the number of clones in the 1st quartile, however the previous data did not indicate the level of antibody expression from the isolated clones or whether this was greater than the control vector when chromatin modifying elements were absent. To answer this question the 24 clones with the highest absorbance values for each construct were expanded into 24 well plates and overgrow cultures set up. Antibody titre was determined by mouse IgG ELISA and results are shown in Figure 4.5 A-D.

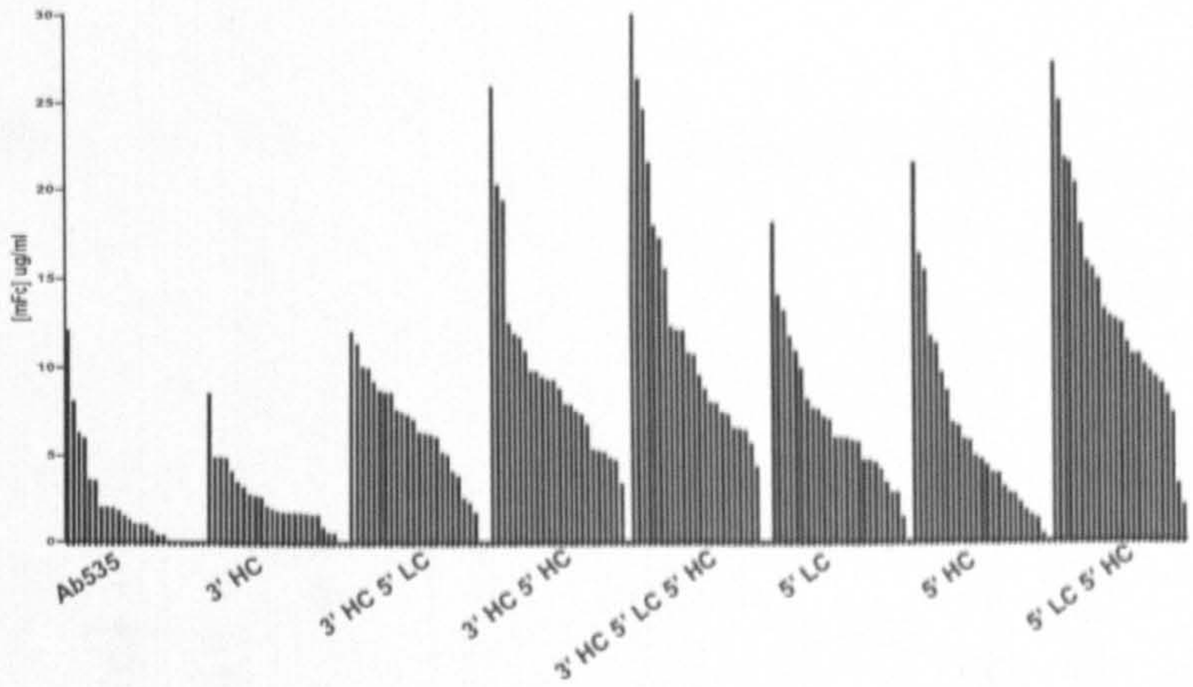
The only A2UCOE-containing construct which did not demonstrate any ability to increase expression was the 3' HC vector. The expression profile of the 24 clones derived from this construct was similar to the profile of the Ab535 clones, (Figure 4.5 A) and there was only a slight increase in the mean expression level of the 24 clones (see Figure 4.6A). The clones with the highest antibody titre were isolated from the cells transfected with the 3' HC 5' LC 5'HC construct with expression levels 2.5-fold higher than the best expressing Ab535 clone (Figure 4.5, 3' HC 5' LC 5' HC). The mean expression level for the 24 clones analysed for each construct was also determined (Figure 4.6A). The highest increase was observed from the cell lines generated using the 5' LC 5' HC vector (6.75-fold increase). This was closely followed by the 3' HC 5' LC 5' HC cell lines where the average expression level was 6.2-fold higher than the Ab535 control clones.

For clones derived from vectors containing the MAR X_S29 element the highest expression levels were observed when the element was at the 3' HC 5' LC locations (Figure 4.5B). Figure 4.6B, illustrates that the average expression from these clones was 2.2 fold higher than the average expression from the Ab535 vector. This result indicates that the 3' HC 5' LC vector performed best. These data were consistent with the pooled stable expression levels and the increase in the number of clones in the 1st quartile.

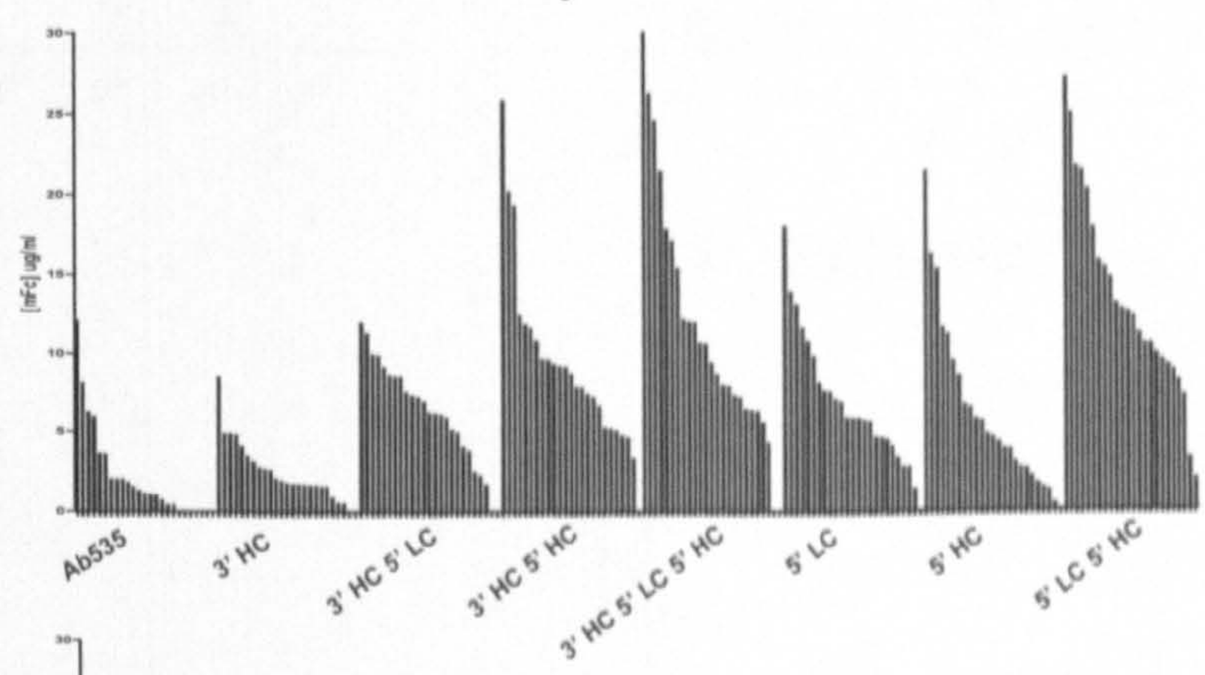
The results from the STAR 40 clones (Figure 4.5C) show that irrespective of the position or number of these elements present in the expression vector, the expression levels observed were similar between expression constructs. Only the 3'HC construct generated any clones with a higher expression level over the Ab535 control. This also resulted in the average expression levels of the cell lines from each expression construct being similar to the control vector (Figure 4.6C). Again these results were consistent with the results observed previously for the pooled stable cell populations.

The highest expression level observed for the cHS4 tandem element was in the cell lines generated using the 3' HC 5' LC 5' HC construct. However, expression levels were similar to those obtained using the Ab535 control vector. In contrast, the highest average expression level is obtained with the 5' LC 5' HC vector (Figure 4.6D).

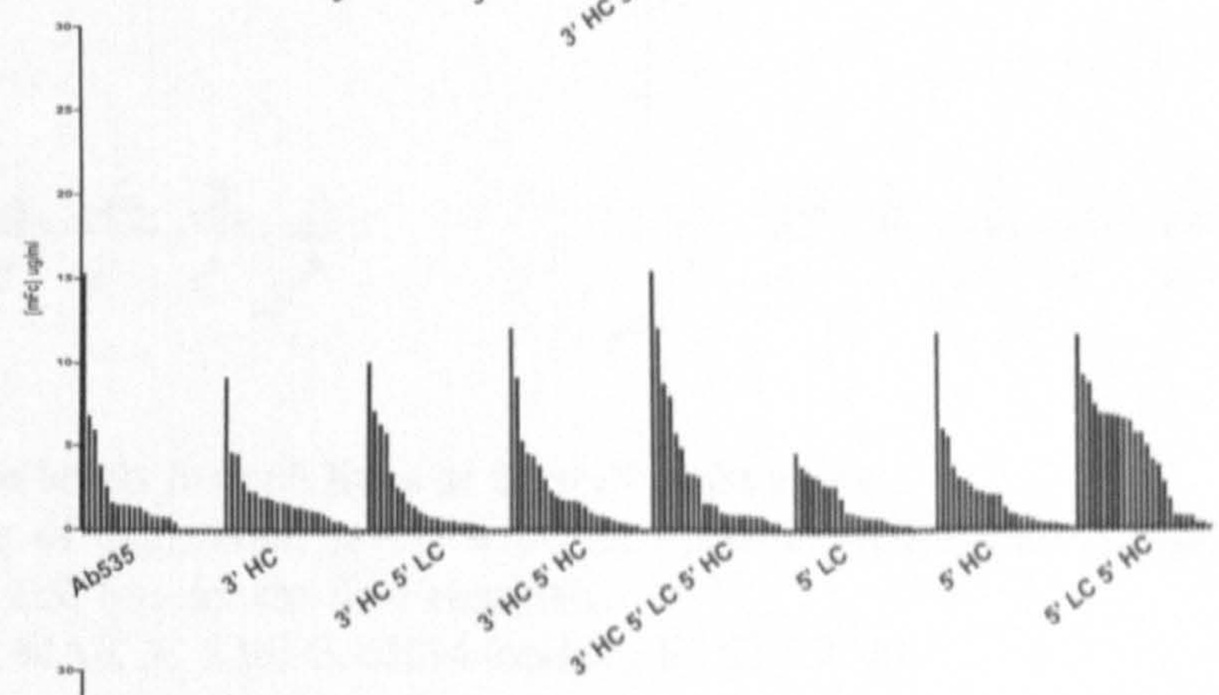
A. 1.5kb A2UCOE



B. MAR X_S29



C. STAR 40



D. cHS4 tandem

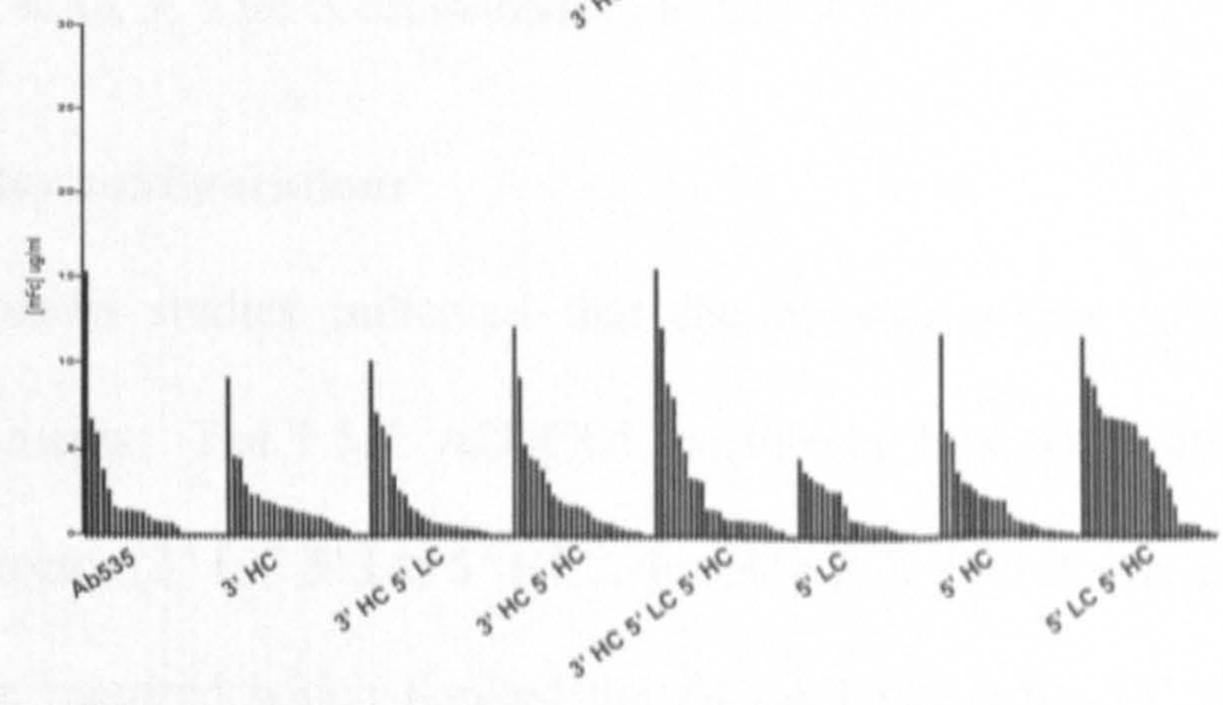


Figure 4.5 Expression levels from overgrown cultures in 24 well plates
Cells from the 24 highest expressing clones at 96 well stage were transferred to 24 well plates and cultured for 14 days. Expression levels were determined by mouse IgG ELISA using MOPC21 as a standard at a starting concentration of 1μg/ml. Samples were added at a 1:50 dilution and titrated 1:5 half way down the plate.

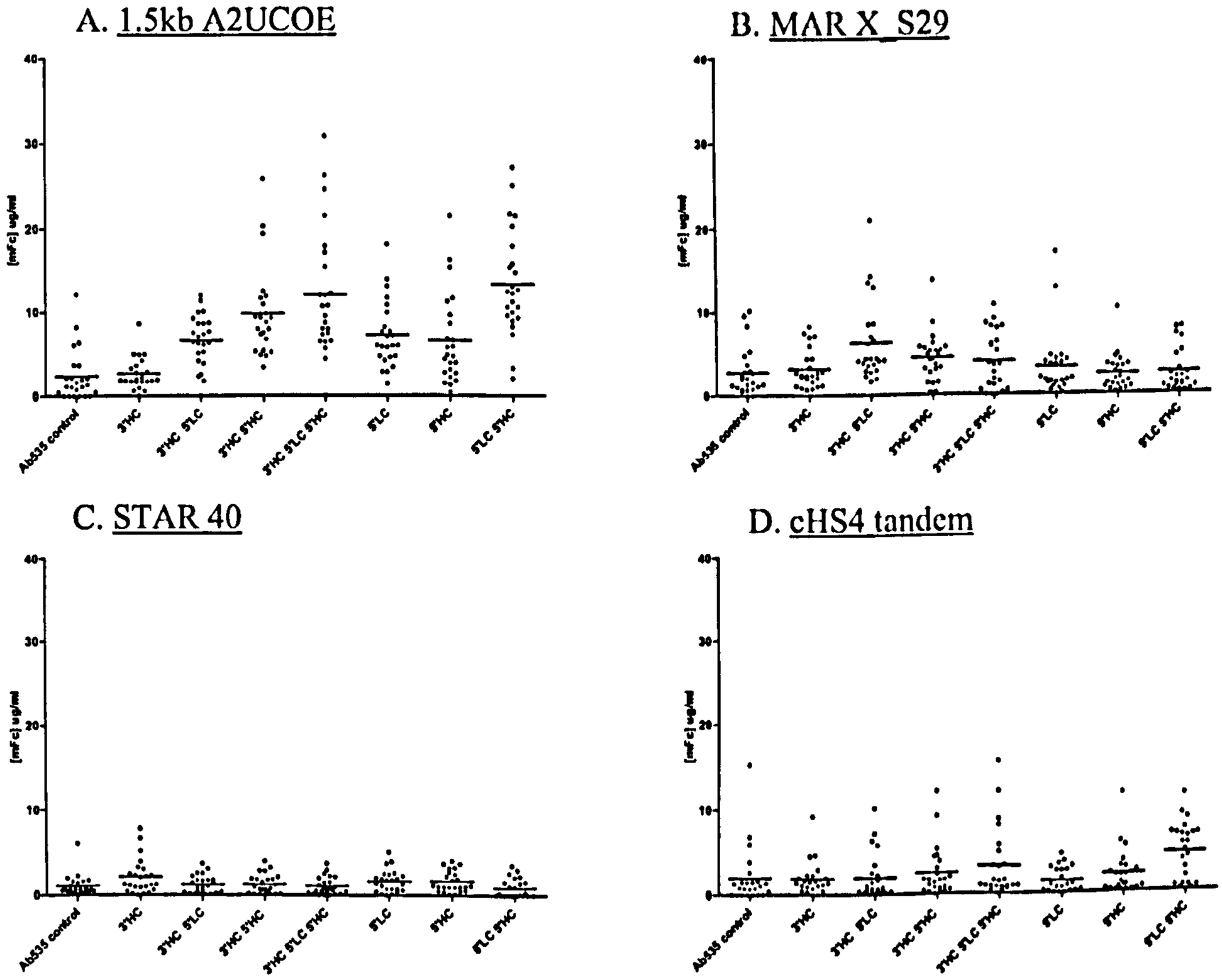


Figure 4.6 Expression levels for cell lines at 24 well plate stage
Graphs showing range of expression levels with the mean expression level highlighted by horizontal bar for each cell line for the four elements.
A. 1.5kb A2UCOE, B. MAR X_S29, C. cHS4 tandem, D. STAR 40.

4.2.3 Optimal vector configurations

The antibody expression studies indicated that the optimal vector configuration differed between elements. The 1.5kb A2UCOE performed best when three copies were present in the vector (3' HC 5' LC 5' HC). For MAR X_S29 it was established that two copies were required which flanked the HC and LC genes (3' HC 5' LC vector). Modest increases in antibody expression were observed when the cHS4 tandem element was located in the vector. Data suggested that the optimal locations for the element were 5' to both the HC and LC (5' LC 5' HC vector). STAR 40

performed poorly in the studies with no vector constructs demonstrating any ability to increase antibody expression. These results reinforced the publication by Otte *et al.* which confirmed the lack of potency of STAR 40 in CHO cells (Otte *et al.* 2007). These data led to the decision to cease investigation of this element and instead focus on the STAR 7 element (Otte *et al.*, 2007). A vector was constructed with the STAR 7 element at the locations described in the paper which was 3' to the HC and 5' to the LC cassettes (see Section 3.2.6.4).

In the initial studies to determine the preferred number and location of the various chromatin modifying elements stable cell lines were established under G418 selection. However, antibody expression was low, <15μg/ml in all the cell lines generated with the exception of cell lines derived from the 1.5 kb A2UCOE containing constructs. In the subsequent studies utilising the optimal vector configurations the selection marker was changed to glutamine synthetase in an attempt to increase the levels of antibody expression. Analysis was again performed on pooled and clonal cell lines.

4.2.4 Comparative analysis of chromatin modifying elements: pooled stable cell line generation under GS selection

The five vector DNAs (four chosen constructs containing the optimal arrangement for each of the chromatin modifying elements, as shown in Table 4.1, and the control Ab535 vector containing no elements), were used to transfect CHO-K1 cells by electroporation after linearisation at the *Pvu*I site within the *ampicillin* resistance gene.

Element	Optimal vector configuration
1.5kb A2 UCOE	3' HC 5' LC 5' HC
MAR X_S29	3' HC 5' LC
STAR 7	3' HC 5' LC
cHS4 tandem	5' LC 5' HC

Table 4.1 Optimal vector configurations utilised in comparative analysis of chromatin modifying elements

Before plasmid DNA can become integrated into the genome it must be linearised by host cell nucleases, and the plasmid may become linearised at any location within the vector which can thus disrupt the transcription units under study. Therefore, in order to standardise this event plasmids were linearised using the unique *Pvu*I site, which cuts outside of the test gene and selection marker gene regions prior to transfection. Following linearisation CHO-K1 cells were transfected as described in Section 2.4.4.2.

The percentage of viable cells and cell number in the pooled stables following the addition of 50μM MSX are shown in Figure 4.7A and B. The percentage of viable cells in the pooled stable generated using the 1.5kb A2UCOE vector remained higher than the Ab535 control with approximately 50% of cells remaining viable in the population. After the viability had reached its lowest point the cell viability immediately increased and by day 14 the cell viability was greater than 80%, this was a 2 fold increase above the number of viable cells in the Ab535 pooled stable at the same time point. The recovery of the 1.5kb A2UCOE pooled stable was 5 days faster than the Ab535 control. The profiles for the MAR X_S29, STAR 7 and cHS4 tandem pooled stable cell lines were similar to the Ab535 pooled stables. The quicker recovery of the 1.5kb UCOE pooled stable may reflect the UCOEs ability to open up

the chromatin environment surrounding the integrated plasmid which will allow more permissive integration events resulting in a higher number of surviving cells in the population.

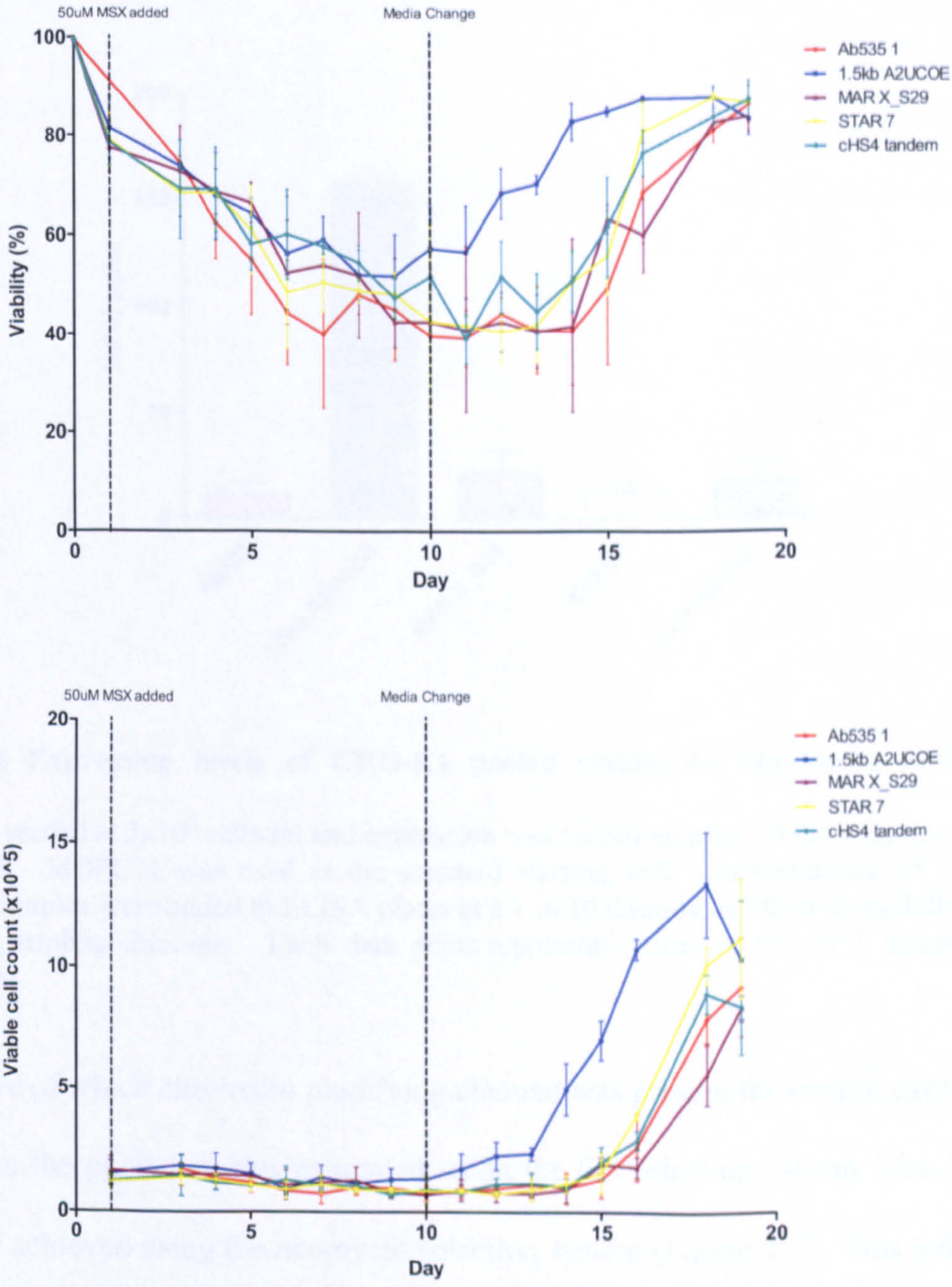


Figure 4.7 Recovery of cell populations in CHO-K1 pooled stable cell lines +/- chromatin modifying elements.

CHO-K1 cells were electroporated and 50μM MSX was added to culture medium 24 hours post transfection. The percentage of viable cells and viable cell numbers in the population was recorded each day. Cell culture medium containing 50μM MSX was replaced on day 10 post transfection.

A. The percentage of viable cells in pooled stables.

B. Viable cell number in pooled stables.

Each line represents the mean +/- SD of 3 independent transfections

Although it is important for the pooled stables to recover quickly, it is also desirable that the cells express the antibody at a high level. Once the cell populations had recovered to greater than 90 % viability antibody expression was assessed in an overgrown culture (Figure 4.8).

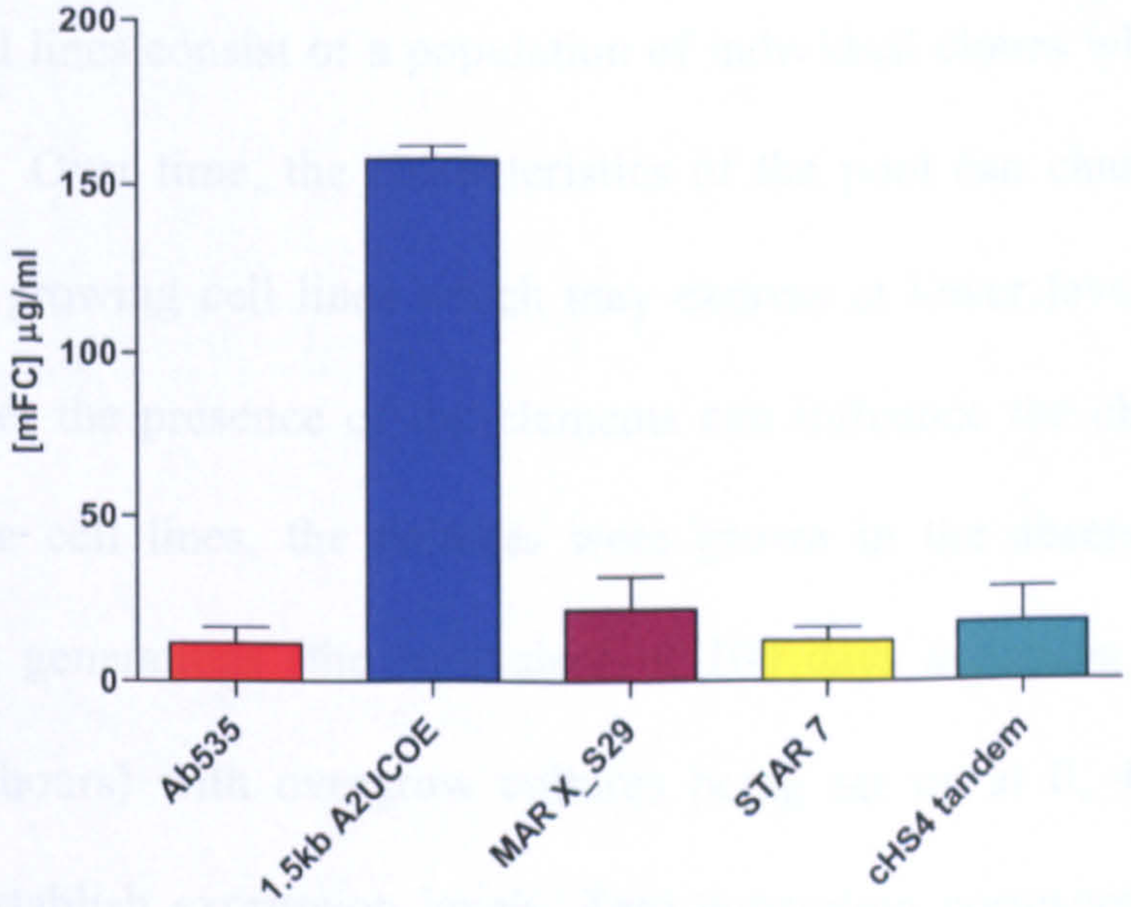


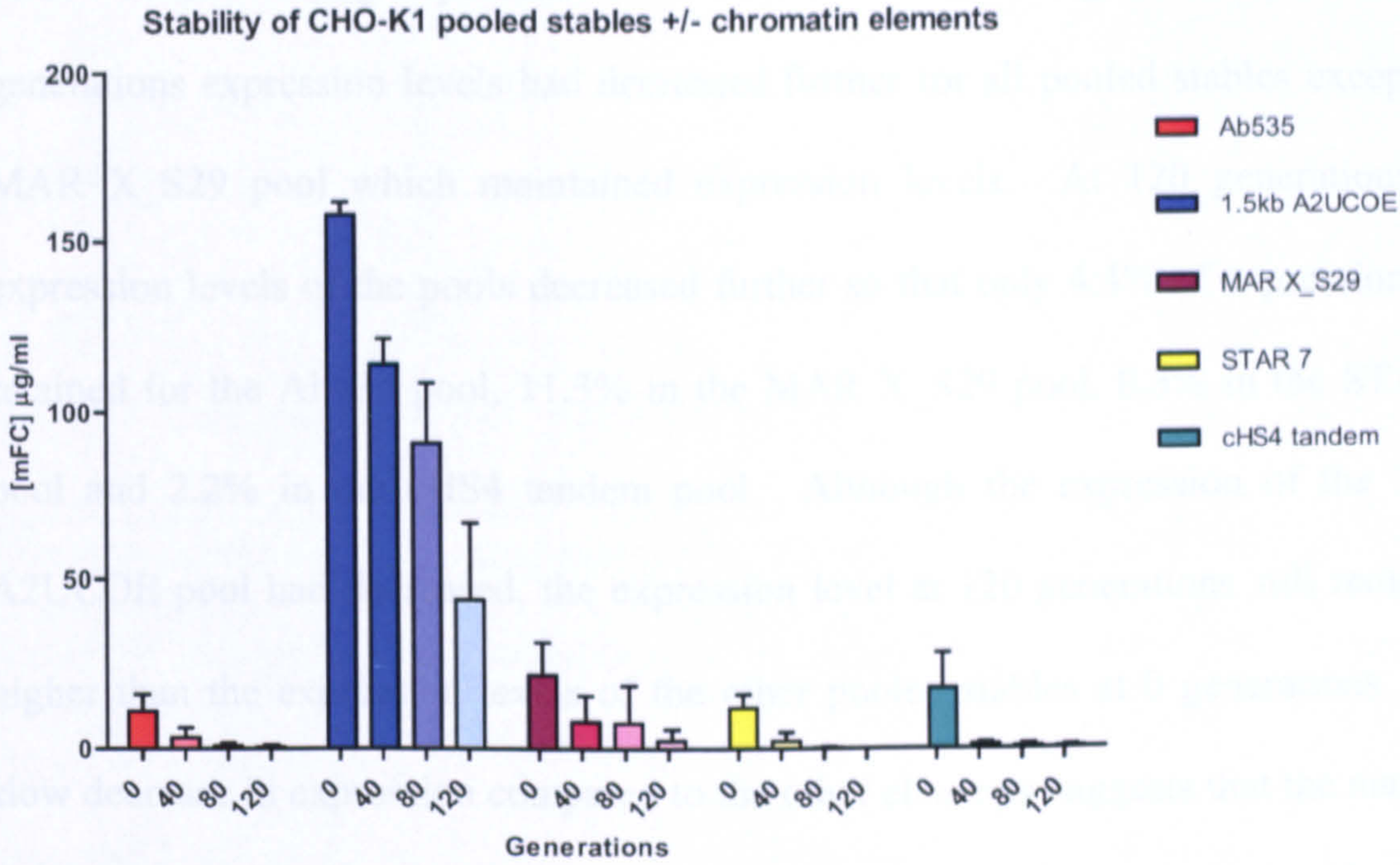
Figure 4.8 Expression levels of CHO-K1 pooled stables +/- chromatin modifying elements
Cells were seeded at $3x10^5$ cells/ml and expression was measured after 10 days using a mouse IgG ELISA. MOPC21 was used as the standard starting with a concentration of 1µg/ml. Triplicate samples were added to ELISA plates at a 1 in 10 dilution and then titrated down the plate using tripling dilutions. Each data point represents mean +/- SD of 3 independent transfections.

Irrespective of which chromatin modifying element was present the overall expression levels from the pooled stables generated using the GS selection system were higher than those achieved using the neomycin selection system (Figure 4.8). This indicated that the GS system was more stringent than the neomycin selection system as cells were only selected which expressed the HC and LC genes at a higher level in contrast to the neomycin selection system which allowed lower expressing clones to survive. The highest expression level observed was in the 1.5kb A2UCOE pooled stable cell line. Expression levels were 15-fold greater than the Ab535 control when the 1.5kb

A2UCOE was present in the vector (Figure 4.8, blue bar). The MAR X_S29 and cHS4 tandem elements showed a modest increase in expression over the control with a 2- and 1.5-fold increase respectively, whereas the STAR 7 element showed no increase in expression above the Ab535 control (Figure 4.8, yellow bar).

Pooled stable cell lines consist of a population of individual clones which all express at varying rates. Over time, the characteristics of the pool can change, due to out growth of faster growing cell lines which may express at lower levels. In order to determine whether the presence of the elements can influence the characteristics of the pooled stable cell lines, the cultures were grown in the absence of selective pressure for 120 generations (the equivalent of 100 days assuming CHO-K1 cells double very 20 hours) with overgrow cultures being set up at 0, 40, 80 and 120 generations to establish expression levels. Zero generation corresponds to the start point of this study, when the pooled stable cell lines had recovered to >90% viability and was equivalent to 18 to 20 generations post-transfection.

A.



B.

	40 Generations	80 Generations	120 Generations
Ab535	30.8	7.9	4.4
1.5kb A2UCOE	72.3	57.6	27.8
MAR X_S29	37.3	34.8	11.3
STAR 7	19.7	2.9	0.3
cHS4 tandem	7.0	4.0	2.2

Figure 4.9 Stability of CHO-K1 pooled stable cell lines
A. CHO-K1 pooled stables +/- chromatin modifying elements were grown in the absence of MSX for 120 generations. At 0, 40, 80 and 120 generations 50ml cultures overgrown cultures with cells seeded at $3x10^5$ cells were overgrown for 10 days. Supernatant was assayed using a mouse IgG ELISA using MOPC21 as a standard at a starting concentration of 1µg/ml. Data represents mean (+/- SD) of 3 independent pooled transfections.
B. Percentage of starting (0 generation) antibody expression remaining at time points measured.

Figure 4.9 shows that in all cases antibody expression in the pooled stables decreased over the 120 generations. Only 30.8% of starting expression remained after 40 generations with the Ab535 pool. There was also a similar decrease in expression during the first 40 generations in the MAR X_S29 pool (37.3% of starting expression level) and a more marked decrease in pools derived from the STAR 7 pool (7% of starting expression) and the cHS4 tandem pool (19% of starting expression). In contrast, expression levels in the 1.5kb A2UCOE pooled stable remained higher with

72.3% of the starting expression level maintained after 40 generations. After 80 generations expression levels had decreased further for all pooled stables except the MAR X_S29 pool which maintained expression levels. At 120 generations the expression levels of the pools decreased further so that only 4.4% of expression was retained for the Ab535 pool, 11.3% in the MAR X_S29 pool, 0.3% in the STAR 7 pool and 2.2% in the cHS4 tandem pool. Although the expression of the 1.5kb A2UCOE pool had decreased, the expression level at 120 generations still remained higher than the expression levels of the other pooled stables at 0 generations. The slow decrease in expression compared to the other elements suggests that the majority of the cells in the 1.5kb A2UCOE pool express antibody, rather than the population of cells being made up of a few cells which expressed antibody to a high level and a large proportion being low or non-expressing. If the population was comprised of the latter a sharp decrease in expression would be expected as the low level, or non expressing cells out grew the cells expressing high levels of antibody as seen with the other elements.

4.2.5 Comparative analysis of chromatin modifying elements: clonal stable cell line generation under GS selection

Clonal cell lines were generated using two methods: limiting dilution and via the Clonepix^{Fl} instrument from previously described pooled stables (Section 4.2.4). In order to generate clonal cell lines by limiting dilution, CHO-K1 cells were transfected as described in Section 2.4.4.2. Cells (4×10^7) were transfected and plated by limiting dilution over forty 96 well plates. Plates were incubated for at least three weeks until colonies were visible. Wells were scored for clonal growth and numbers are shown in Table 4.2.

Cell Line	Single Colonies
Ab535	78
1.5kb A2UCOE	123
MAR X_S29	53
STAR 7	81
cHS4 tandem	60

Table 4.2 Total number of wells with single colonies

The highest number of wells with single colonies was achieved when the 1.5kb A2UCOE was present in the expression vector. The presence of the other elements in the vectors did not appear to have a beneficial effect on the number of colonies produced.

For the derivation of clonal cell lines from the pooled stable cell lines, single colonies were isolated using the Clonepix^{Fl} instrument. Cell lines were plated in semi solid CloneMedia and 10 days after plating colonies were picked using the Clonepix^{FL} instrument under white light. A total of 192 colonies for each construct were picked into two 96 well plates.

To assess the distribution of expressing cells in the clonal cell lines isolated using the two methods spot ELISA was perfomed as described in Section 4.2.2. The percentage of clones in each quartile was calculated as described in Figure 4.3. Results from clones isolated by limiting dilution and growth in CloneMedia are shown in Figure 4.10 and 4.11 respectively.

A.

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
Ab535	19.2	6.4	2.6	71.8
1.5kb A2UCOE	78.9	1.6	1.6	17.9
MAR X S_29	41.5	9.4	3.8	45.3
STAR 7	21.0	3.7	7.4	67.9
cHS4 tandem	38.3	1.7	1.7	58.3

B.

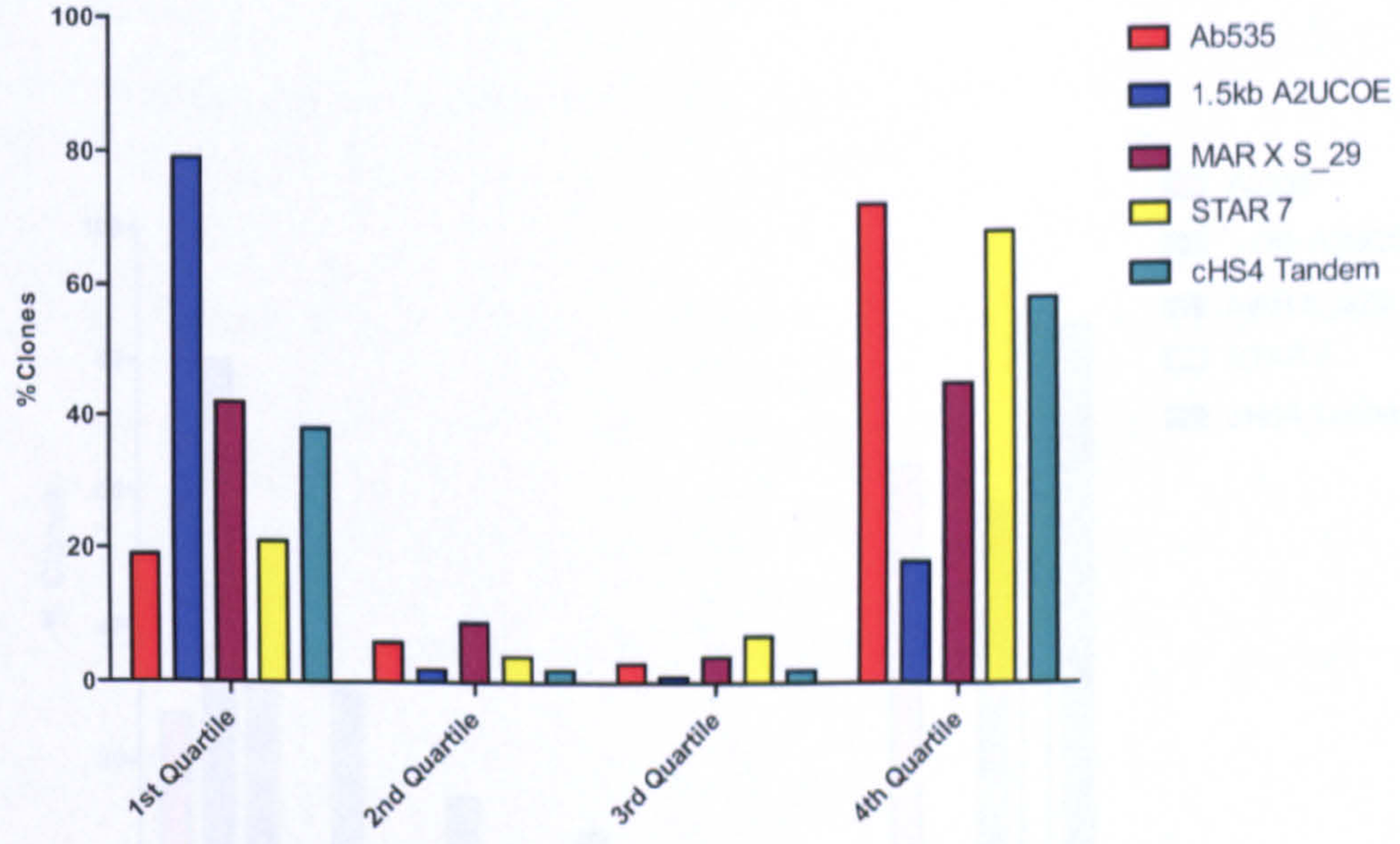


Figure 4.10 Distribution of expressing clones isolated by limiting dilution
Supernatant from wells containing single colonies was assayed using a spot ELISA with samples added to the plate at a 1:40 dilution. A. Table of results comparing the percentage of clones in each quartile for each element. B. Graph showing % of clones in each quartile.

A.

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
Ab535	27.4	6.1	2.4	64.0
1.5kb A2UCOE	80.3	15.2	3.4	1.1
MAR X_S29	31.6	7.5	2.7	58.3
STAR 7	7.3	2.7	2.7	87.3
cHS4 tandem	31.4	10.1	4.7	85.2

B.

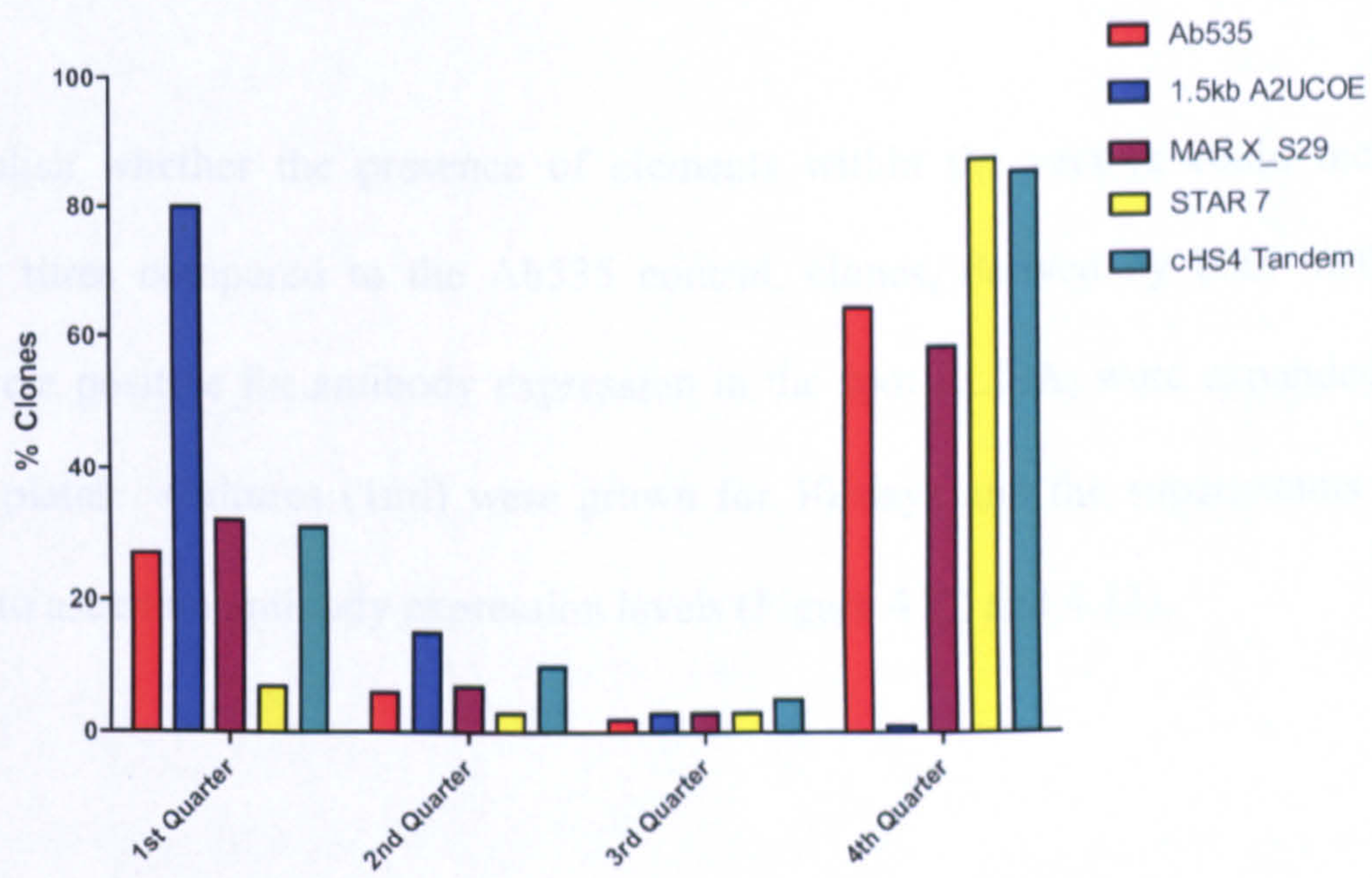


Figure 4.11 Distribution of expressing clones isolated from GS pooled stables
Colonies picked from CloneMedia were assayed for expression from growth in 96 well plates. Supernatant was assayed using a spot ELISA with samples added onto the plate at a 1:40 dilution. A. Percentage of cells within each quartile. B. Graph showing % of clones in each quartile.

The profile of the distribution of clones for the two methods were similar.

Irrespective of the method used to isolate the clonal cell lines the presence of the 1.5kb A2UCOE in the expression vector shifted the distribution of expressing clones so that the majority were located in the first (highest expressing) quartile (79% from limiting dilution and 80% from pooled stable). The presence of the MAR X_S29 and the cHS4 tandem element in the vector produced a 2-fold increase in the number of clones in the first quartile compared to the control vector when clones were isolated

using limiting dilution. In contrast, when clonal cell lines were isolated from the pooled stables this increase in the percentage of clones in the 1st quartile was not observed.

The distribution of expressing clones derived from the STAR 7 containing construct isolated using limiting dilution was similar to the Ab535 vector control. However, the percentage of clones in the 1st quartile isolated from the STAR 7 pooled stable was much lower than the Ab535 control with only 7% of clones in this quartile.

To establish whether the presence of elements within the vectors could increase antibody titres compared to the Ab535 control, clones, derived by both methods, which were positive for antibody expression in the spot ELISA, were expanded into 24 well plates. Cultures (1ml) were grown for 10 days and the supernatants were assayed to ascertain antibody expression levels (Figure 4.12 and 4.13).

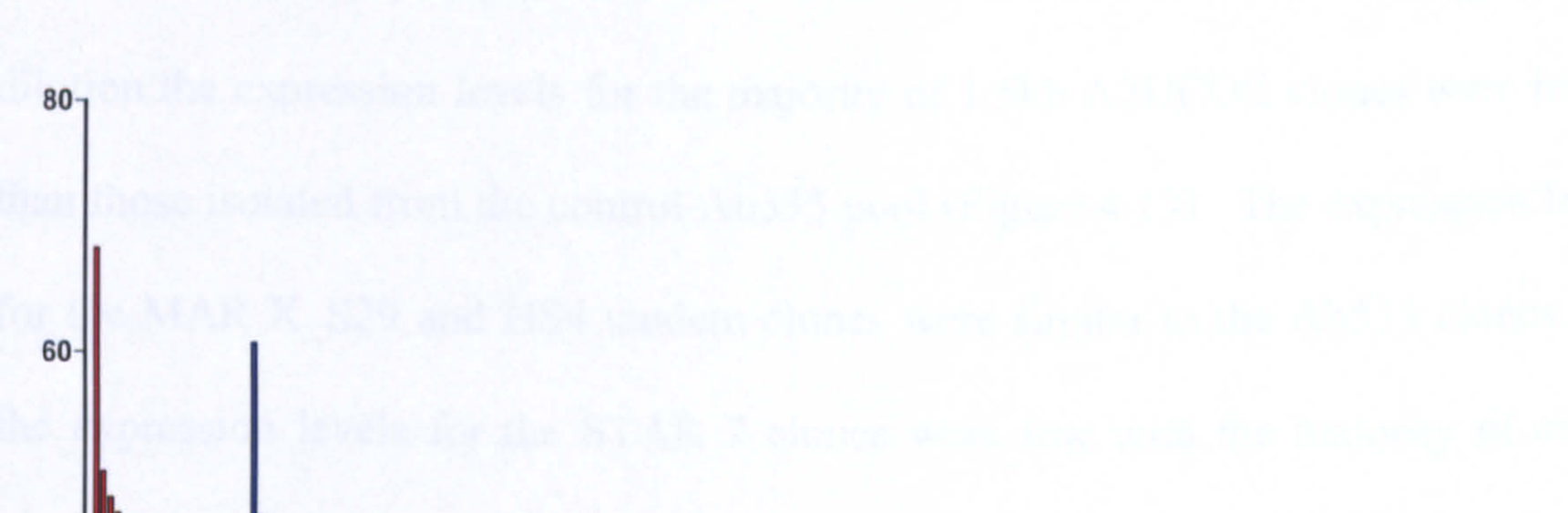


Figure 4.12 Expression levels of clones isolated by limiting dilution +/- chromatin modifying elements in 24 well plates

All clones positive for expression in the spot ELISA were expanded into 24 well plates and grown for 10 days. Supernatant was assayed using a mouse IgG ELISA with MOPC21 as a standard. Supernatant was added at a 1:50 dilution and titrated half way down the plate using 1:5 dilutions

For clones generated using limiting dilution the highest expressing clone was isolated from the Ab535 cell line. This had an antibody titre of over 60μg/ml. The expression levels of the remaining Ab535 clones ranged from 50μg/ml to 5μg/ml. The expression levels of the clones generated using the 1.5kb A2UCOE vector showed a similar range with antibody titres ranging from 60μg/ml to under 1μg/ml. The expression levels of the MAR X_S29 clones were lower, with antibody titres ranging from 39μg/ml to 4μg/ml. Expression levels from STAR 7 and HS4 tandem clones were the lowest with the best clones only expressing 15μg/ml and 17μg/ml respectively.

For clones isolated from the pooled stables only the top 48 clones were expanded to 24 well plates with the exception of the STAR 7 cell line where only 18 colonies were positive for antibody expression. In contrast to the clones isolated using limiting dilution the expression levels for the majority of 1.5kb A2UCOE clones were higher than those isolated from the control Ab535 pool (Figure 4.13). The expression levels for the MAR X_S29 and HS4 tandem clones were similar to the Ab535 clones, and the expression levels for the STAR 7 clones were low with the majority of clones expressing below 10μg/ml.

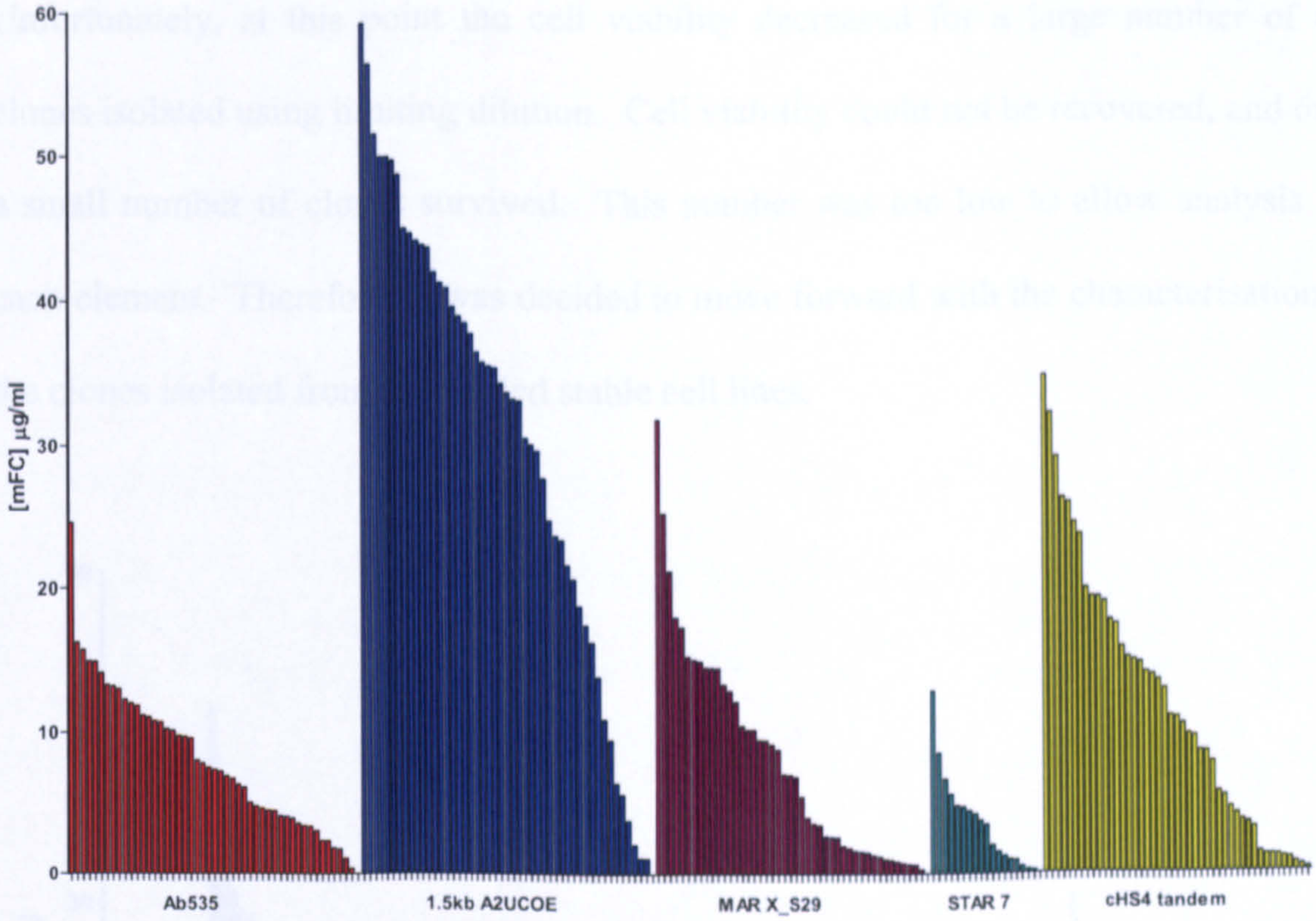


Figure 4.13 Expression levels of clones isolated from pooled stables +/- chromatin modifying elements in 24 well plates

Cells were expanded into 24 well plates and grown for 10 days. Supernatant was assayed using a mouse IgG ELISA with MOPC21 as a standard at a starting concentration of 1µg/ml.

Clones isolated from limiting dilution were cultured in 6 well plates. Several clones failed to grow, however overgrow cultures were set up with the remaining clones. The larger culture volume enabled cell numbers to be determined and therefore cells were seeded at a known cell density, ($3x10^5$) cells/ml, and cultured for 10 days. Supernatants were assayed by mouse IgG ELISA and the results can be seen in Figure 4.14. In contrast to the results at the 24 well stage the highest expression levels were from 1.5kb A2UCOE clones. Although the highest antibody titre was seen with the 1.5kb A2UCOE, there was still a wide range of expression levels when the UCOE was in the vector. The range of expression levels seen in the clones isolated from the other elements was similar to the Ab535 control.

Unfortunately, at this point the cell viability decreased for a large number of the clones isolated using limiting dilution. Cell viability could not be recovered, and only a small number of clones survived. This number was too low to allow analysis for each element. Therefore, it was decided to move forward with the characterisation of the clones isolated from the pooled stable cell lines.

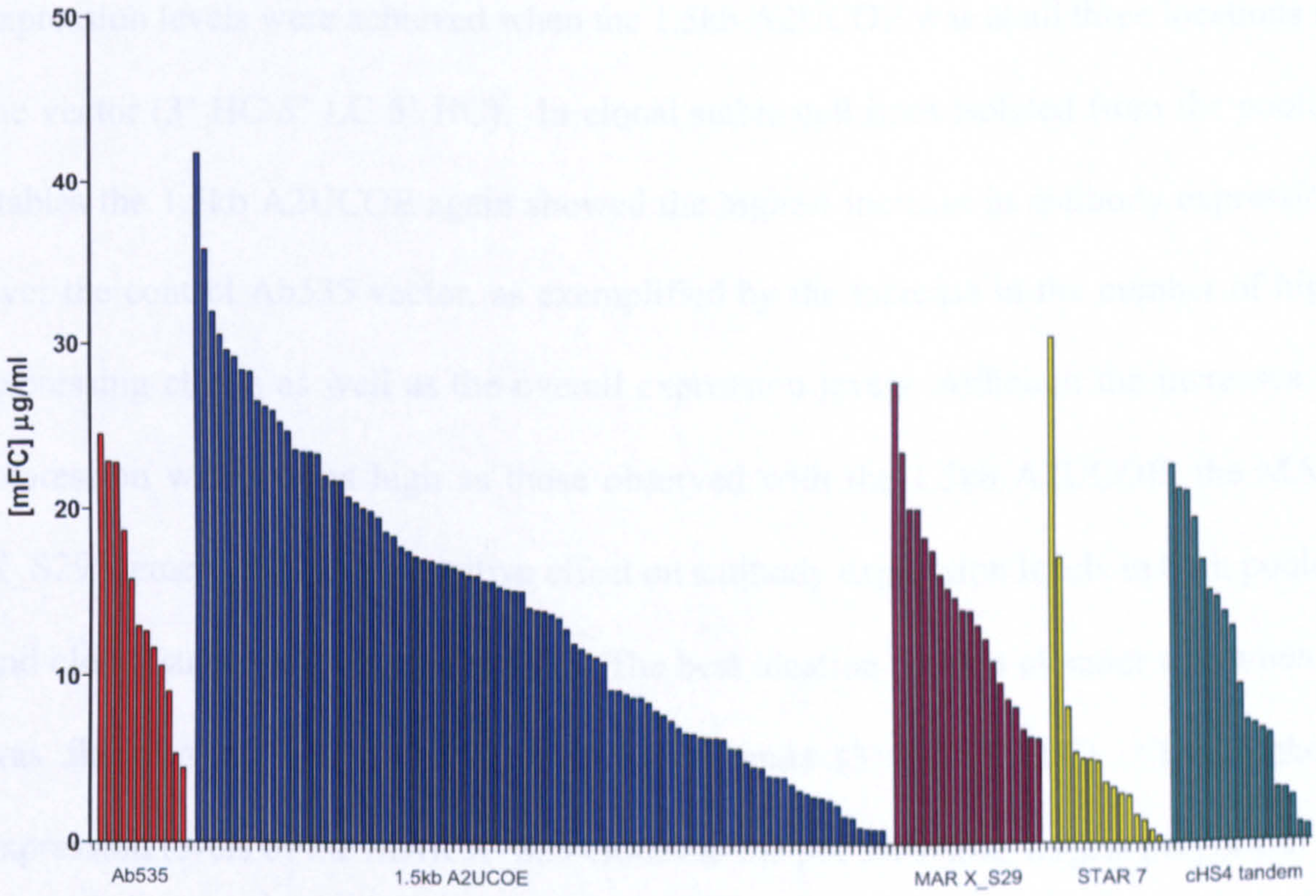


Figure 4.14 Expression levels of clones +/- elements in 6 well plates
Cells were seeded at 3 x 10^5 cells/ml in 3 ml media and grown for 10 days. Supernatant was assayed using a mouse IgG ELISA with MOPC21 as a standard at a starting concentration of 1µg/ml. Samples were added at a 1:10 dilution and titrated down the plate at a 1:3 dilution.

4.3 Discussion

The aim of the work described in this Chapter was to assess the effect of various chromatin modifying elements on antibody expression in stable cell lines. Initial experiments were aimed at establishing the optimal locations and combination of the

individual elements. The results showed that the chromatin modifying elements had an effect on antibody expression in both pooled and clonal stable cell line generation and different combinations performed best for each element. It was also observed that the elements affect antibody expression to varying degrees.

In the initial pooled stable cell lines, selected using G418, the highest antibody expression levels were achieved when the 1.5kb A2UCOE was at all three locations in the vector (3' HC 5' LC 5' HC). In clonal stable cell lines isolated from the pooled stables the 1.5kb A2UCOE again showed the highest increase in antibody expression over the control Ab535 vector, as exemplified by the increase in the number of high expressing clones as well as the overall expression levels. Although the increases in expression were not as high as those observed with the 1.5kb A2UCOE, the MAR X_S29 element did have a positive effect on antibody expression levels in both pooled and clonal stable cell line generation. The best location for this element was when it was flanking the HC and LC transcription units (3' HC 5' LC). The highest expression levels of the MAR X_S29 clones in the pooled stable, largest proportion of expressing clones and highest expression levels at the 24 well plate stage were all observed in the cell lines produced with this vector. Previously, data presented by Otte *et al.* (2007) suggested that the incorporation of MAR X_S29 gave a marked increase in expression levels, compared to clones derived using an expression vector containing UCOEs. However, data generated from the work described in this Thesis indicated that MAR X_S29 was not as potent as the A2UCOE in increasing expression levels. Since the work described in this Thesis was started it became apparent that MAR X_S29 had limited activity compared to the other human MAR elements identified in the bioinformatic search of the human genome. It is said to

perform better in a mesoangioblastic cell line (Mermod, 2007) and a different MAR element, MAR 1_68, has better activity in CHO cells (Girod *et al.*, 2007).

The results for the cHS4 tandem and STAR 40 elements were less convincing with only small increases seen over the Ab535 control. For the cHS4 tandem element this may be due to its proposed functionality. It is classed as an insulator and in contrast to the A2UCOE cannot 'open up' heterochromatin (Chung *et al.* 1993; Antoniou *et al.*, 2003). This could explain why the number of expressing clones and overall expression levels were similar to the Ab535 control. The data presented is also consistent with a previous study in which it was shown that this element could not increase enhanced green fluorescent protein (EGFP) expression levels in clonal stable cell line generated in CHO cells (Otte *et al.*, 2007). The STAR 40 vectors also gave only modest increases in expression compared to the control Ab535 vector. These results were inconsistent between pooled and clonal stable cell lines with no vector configuration being preferable. Whilst this investigation was in progress, a publication by Otte and colleagues has described the lack of potency of the STAR 40 element in CHO cells (Otte *et al.*, 2007). The disappointing results observed with this element were therefore not unexpected and led to the use of the STAR 7 element, which was described as being more potent in CHO cells. However, in the expression vector constructs described in this Thesis this element also gave poor results with no improvements in number of clones expressing antibody and levels of expression. Although STAR 7 is described as being more potent than STAR 40 in CHO cells, the data presented by Otte *et al.* (2007) shows STAR 7 being used in combination with another STAR element, STAR 67. It was demonstrated that this STAR element had a low level of anti repressor activity but it could have a positive, additive effect on

antibody expression when placed upstream of a promoter driving transgene expression (Otte *et al.*, 2007; Van Blokland *et al.*, 2007). This suggests that STAR 67 may function differently to the other STAR elements and is an enhancer of expression rather than a true anti-repressor element. Therefore, on its own, STAR 7 may not be able to enhance transgene expression.

Data generated using the GS selection marker confirmed that the presence of the 1.5kb A2UCOE in the vector resulted in the highest levels of antibody expression in the pooled stables. The results observed are consistent with those obtained in a previous study where expression levels were increased six-fold when a 3.2kb UCOE (from the 5' untranslated region of a murine ribosomal protein) were present in the expression vector compared to the control vector (Ye *et al.*, 2010).

In media containing MSX cell death and recovery followed a similar pattern for all of the cell lines with the exception of those derived from the 1.5 kb A2UCOE. In this case viability remained higher and recovery was markedly quicker. Presumably, the ability of UCOEs to 'open up' chromatin allows a greater number of permissive integration events to occur, enabling efficient transcription of GS. Therefore a greater number of cells can survive the selection pressure. However, due to the vector configuration after linearisation the GS transcription unit does not have a UCOE upstream of its promoter (to function correctly the UCOE should be immediately upstream of the promoter driving expression of the transgene, Williams *et al.*, 2005) which indicates that UCOEs may possibly be able to provide benefits to genes that are not operably-linked.

Irrespective of the method utilised to isolate single colonies the presence of the UCOE in the expression vector increased the proportion of expressing clones, which is consistent with data published previously (Benton *et al.*, 2002; Williams *et al.*, 2005; Ye *et al.*, 2010). Interestingly, when clonal cell lines were generated by limiting dilution the increase in expression in the 1.5kb A2UCOE containing clones over the clones derived using the Ab535 control vector was less evident than in the pooled stables. More clones were isolated that expressed antibody but none of these had a higher productivity than the clones derived using the Ab535 control vector in overgrown cultures in 24 well plates. In contrast, clonal cell lines isolated from the GS pooled stables correlated with data generated when determining the optimal vector configurations, where a large number of the 1.5kb A2UCOE clones had higher expression levels than the clones isolated from the Ab535 control. In two of the published studies (Benton *et al.*, 2002; William *et al.*, 2005) clones were derived from pooled stables and expression levels of clones generated using an expression vector containing UCOEs were higher than those derived from a non-UCOE expression vector which is consistent with the findings presented in this Thesis. In the study by Ye *et al.*, 2010 single colonies were derived using limiting dilution. Although the data presented is not conclusive a number of non-UCOE clones expressed at the upper detection limit of the assay indicating that similar expression levels could be achieved, which again, is also in agreement with data presented in this Thesis.

A possible explanation for the increased expression in the clones derived from the 1.5kb A2UCOE pooled stable is that there may be a lack of diversity in the pool of stably transfected cells generated using the Ab535 control vector. After the addition of selection pressure the pool of cells were allowed to recover and during this time

clones expressing antibody at a high level may have been lost from the population due to the outgrowth of faster growing, low expressing cells. In contrast, the 1.5kb A2UCOE pooled stable contained more cells with higher expression levels. Therefore, a greater number of clones expressing at a high level could be isolated. When colonies were derived using limiting dilution cells were plated out immediately after the addition of MSX. This meant that, although it was a rarer event, clones could still be isolated from cells transfected with the Ab535 control vector which expressed at a level comparable to clones derived using the 1.5kb A2UCOE vector.

The work described in this Chapter has established the effects that the various chromatin modifying elements were having on antibody expression in stable cell lines. To investigate how they were potentially functioning within the cell a number of clones were chosen to characterise further and this work is described in the subsequent Chapters.

CHAPTER 5

Molecular Genetic Characterisation Of Stably Transfected Cell Lines

5.1 Introduction

Having established the optimal arrangement of the different chromatin modifying elements in the antibody expression vectors, clonal cell lines were generated employing the preferred constructs using the GS selection system. The clonal cell lines were selected from pools of stably transfected cells following selection with MSX. This chapter will describe the molecular genetic characterisation of these clonal cell lines in order to obtain a degree of insight into how chromatin modifying elements effect antibody expression levels.

The expression level of a stably expressed mAb can be dependent upon a number of factors that includes transgene copy number and site of integration.

Identifying molecular differences or similarities between clones and/or elements will should provide an improved understanding of how the elements work within CHO cells.

A total of ten clones for each of the five test constructs, (Ab535 control, 1.5kb A2UCOE, MAR X_S29, STAR 7 and cHS4 tandem) were chosen for further analysis. HC and LC transgene copy number and integrity was assessed using TaqMan Q-PCR and Southern blot hybridisation.

The principle upon which the probe based TaqMan Q-PCR is as follows. A small region of the integrated plasmid is amplified with forward and reverse oligonucleotides. A fluorogenic probe that anneals within the amplified region is also included. The probe contains a fluorescent dye at the 5' end and a quencher dye at the

3' end. As amplification of the DNA takes place the quencher is removed from the probe due to the 5' exonuclease activity of the polymerase. The amount of fluorescence produced is proportional to the amount of target DNA present. In Q-PCR the fluorescence generated following degradation of the probe is plotted over time and the point at which the fluorescence exceeds the background level is given a value, which is referred to as the cycle threshold (C_T). Samples with a higher level of target DNA will have a lower C_T value as the fluorescence appears after fewer cycles of PCR amplification.

Southern blotting is a technique used for isolating and identifying specific fragments of DNA (Southern *et al.*, 1975). The mixture of fragments is subjected to electrophoresis through an agarose gel, followed by denaturation to form single stranded fragments. These are transferred or 'blotted' onto a filter membrane which immobilises them in their relative positions. Labelled probes, which are specific for a certain region are then added. These hybridise with any complementary fragments on the filter, which can then be detected.

When Southern blotting was first described radioactive probes were utilised. More recently alternative methods for probe labelling have been used (Rihn *et al.*, 1995; Solanas and Escrich, 1997). An example of this is Digoxigenin (DIG). The DIG-labelling method is based on a steroid isolated from digitalis plants (*Digitalis purpurea* and *Digitalis lanata*). Digoxigenin is linked to the C-5 position of uridine nucleotides via a spacer arm containing eleven carbon atoms which forms DIG-11-dUTP. This can be incorporated into the DNA probe by enzymatic nucleic acid synthesis using DNA polymerases. The DIG labelled probe can then be detected either colourmetrically or by chemiluminescent detection.

Southern blotting can also be used to determine clonal diversity. In this study a probe was designed which hybridised to the 3' end of the linearised integrated plasmid (so called 'end-fragment analysis'), which enabled clones with different integration sites to be distinguished due to the difference in the size of the fragment visualised by probe hybridisation.

Chromosomal integration sites of the plasmid DNA was determined by fluorescence in situ hybridisation (FISH). This was performed on three clones containing 1.5kb A2UCOE transgenes. FISH is a cytogenic method that uses fluorescent probes to detect and localize the presence or absence of specific DNA sequences on chromosomes. Probes will only bind to those parts of the chromosome where there is a high degree of sequence similarity. The fluorescent probes bound to the chromosomes can then be visualised using fluorescent microscopy.

5.2 Results

5.2.1 Plasmid copy number determination by TaqMan Q-PCR

HC and LC copy numbers were determined using TaqMan Q-PCR as described in Section 2.3.2. Oligonucleotides were designed (shown in Appendix 2) that amplified a region of DNA in the murine γ1 full length and Kappa constant regions as illustrated in Figure 5.1.

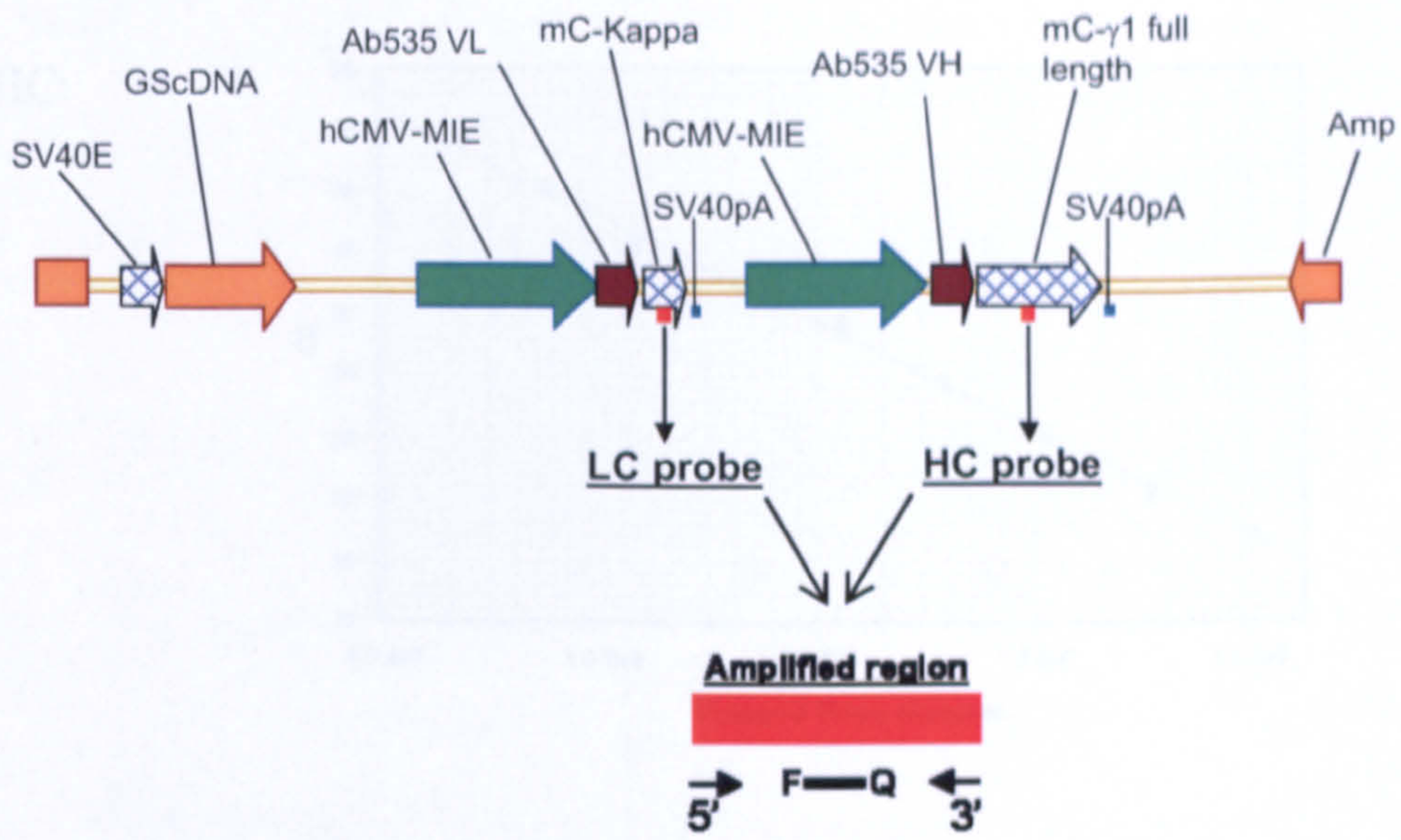
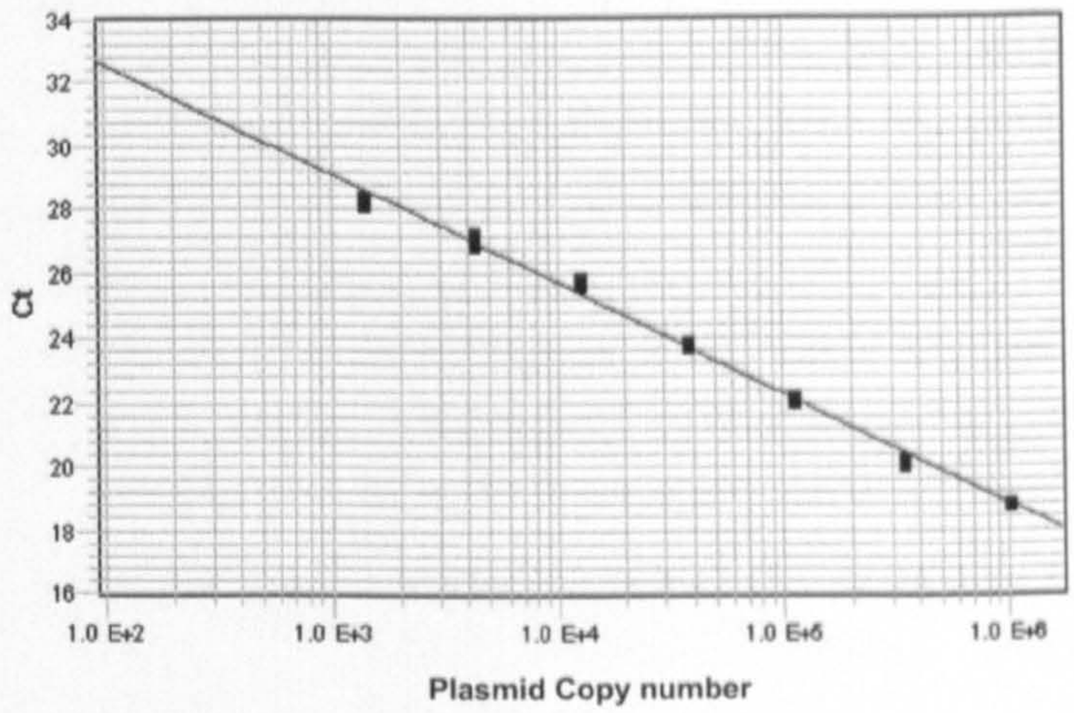


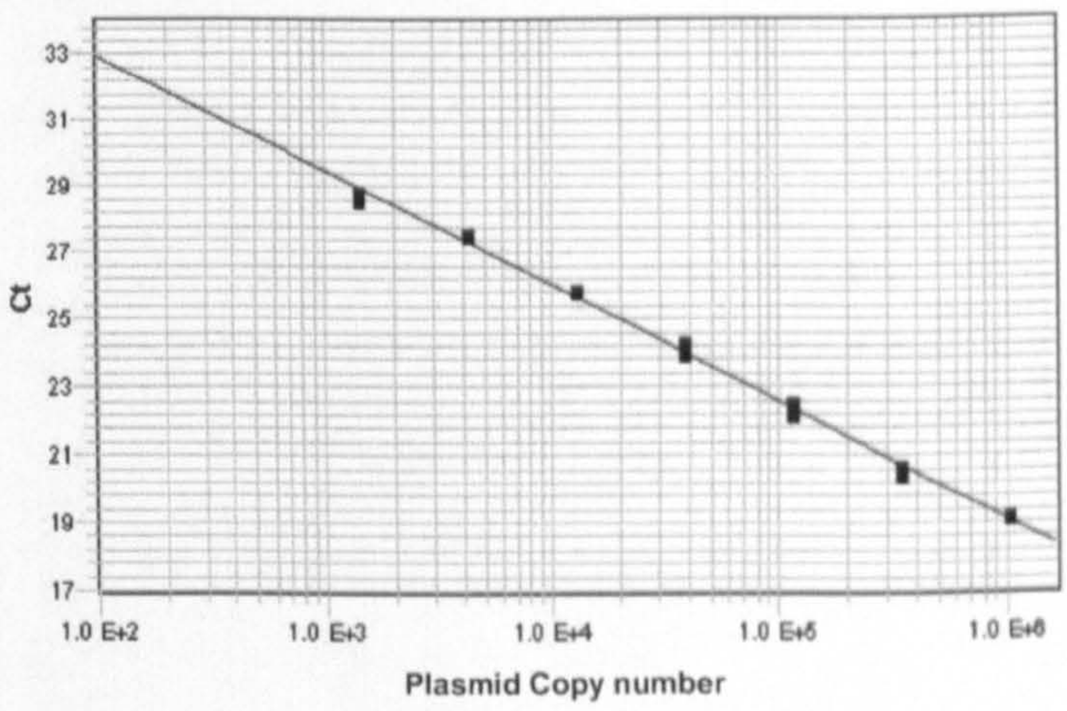
Figure 5.1 Q-PCR oligonucelotide and probe locations
Map of linearised expression plasmid highlighting location of oligonucleotides and probes designed to anneal in the Kappa and γ1 full length constant regions. Annealing site indicated by red squares. Amplified region expanded to highlight 5' and 3' oligonucleotides and probe containing fluorescence (F) at 5' end and quencher (Q) at 3' end.

In order to standardise the amount of input gDNA used in the Q-PCR reactions, quantification of hamster GAPDH was used as an internal control. Oligonucleotides (3.1 to 3.3, shown in Appendix 2) were designed to amplify a region of the GAPDH gene and a standard curve was generated using CHO-K1 gDNA. Standard curves for HC and LC DNA were also generated using a serial dilution of known copy numbers of plasmid DNA diluted within CHO-K1 gDNA (see Section 2.3.2.1). Examples of the standard curves generated can be seen in Figure 5.2. Together, these curves were used to calculate the number of cells in the input gDNA samples (see Table 5.1), and calculate HC (see Table 5.2) and LC (see Table 5.3) transgene copy number using the absolute quantification software (Applied Biosystems). An example of the data generated for the ten MAR transgene containing clones is shown in Tables 5.1 to 5.3. The remaining data is shown in Appendix 4.

HC:



LC:



GAPDH:

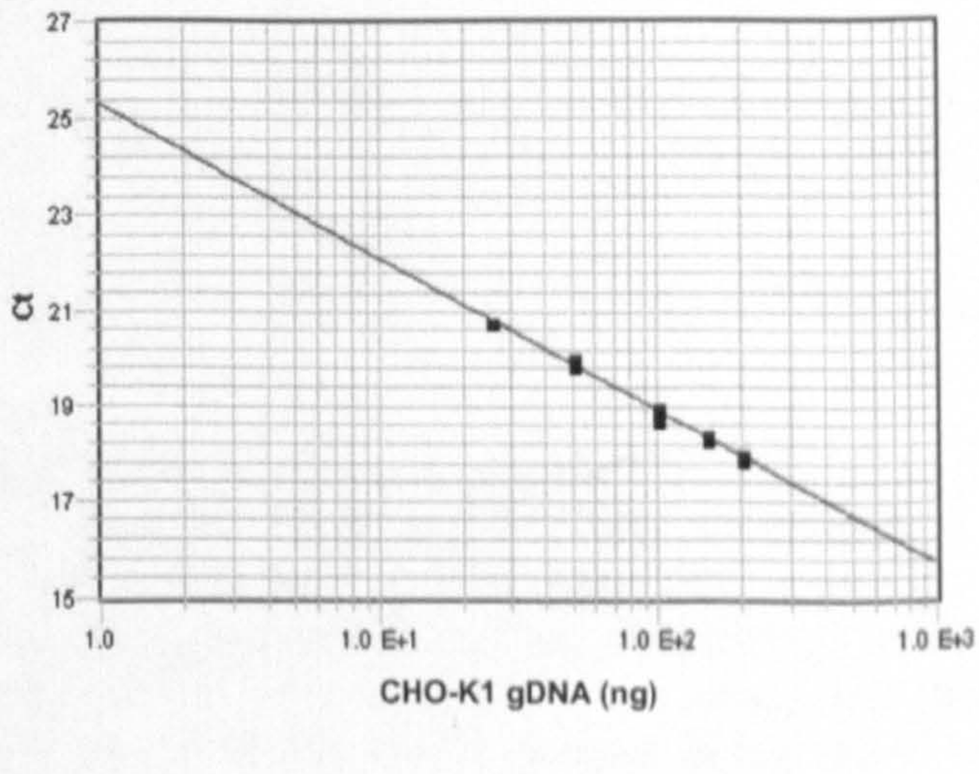


Figure 5.2 Standard curves generated for Q-PCR
Standard curves for HC and LC DNA were generated by a 1:3 serial dilution of $1x10^6$ copies of plasmid DNA (sampled in triplicate) and plotting quantity of plasmid DNA against the corresponding C_T value. A standard curve was generated for GAPDH by plotting known quantities of CHO-K1 gDNA against the corresponding C_T value for triplicate samples.

Clone	C_T	Quantity (ng)	Avg quantity (ng)	# cells
M1	20.63	26.99	27.96	4236
	20.56	28.54		
	20.57	28.34		
M2	21.11	18.96	20.41	3093
	20.82	23.55		
	21.12	18.74		
M3	18.41	140.32	138.34	20961
	18.41	140.62		
	18.47	134.09		
M4	18.89	98.12	98.02	14852
	18.78	107.13		
	19.03	88.81		
M5	18.96	93.17	109.65	16613
	18.62	119.78		
	18.67	115.99		
M6	18.54	127.35	120.51	18259
	18.76	108.36		
	18.56	125.82		
M7	19.84	48.75	47.59	7211
	19.85	48.16		
	19.92	45.87		
M8	20.30	34.44	36.31	5501
	20.19	37.62		
	20.21	36.88		
M9	18.73	111.03	123.08	18648
	18.54	127.17		
	18.50	131.04		
M10	18.54	127.10	122.15	18508
	18.62	119.96		
	18.63	119.40		

Table 5.1 Calculating number of cells per sample
An example generated using the MAR X_S29 clones showing how the GAPDH standard curve was used to standardise input gDNA quantities. Average quantities (ng) were calculated in the Absolute quantification (Applied Biosystems) software using the GAPDH standard curve. Number of cells were calculated by assuming that the DNA mass of the cell is approximately 6.6pg (based on the rodent genome being $3x10^9$ bp in length; Marshall, 2001)

HC:

Clone	C_T	Quantity	Qty Mean	Qty SD	Copy per cell	Copy per cell SD
M1	25.00	16542.47	15909.64	558.54	3.76	0.13
	25.07	15700.96				
	25.09	15485.48				
M2	25.97	8402.46	9069.18	658.51	2.93	0.21
	25.76	9719.17				
	25.86	9085.91				
M3	22.93	69782.29	59646.21	10306.97	2.85	0.49
	23.14	59979.88				
	23.43	49176.45				
M4	23.46	48185.74	45379.68	2457.59	3.06	0.17
	23.60	43610.23				
	23.58	44343.07				
M5	24.74	19827.11	20548.71	633.26	1.24	0.04
	24.65	21011.93				
	24.67	20807.10				
M6	22.96	67996.88	60049.11	7865.65	3.29	0.43
	23.34	52268.24				
	23.15	59882.23				
M7	25.07	15744.32	14648.79	1178.85	2.03	0.16
	25.16	14800.70				
	25.30	13401.35				
M8	25.09	15443.34	15162.93	529.53	2.76	0.10
	25.18	14552.17				
	25.09	15493.29				
M9	23.92	34880.91	40215.29	4776.49	2.16	0.26
	23.67	41668.76				
	23.59	44096.19				
M10	24.48	23710.95	22796.66	997.56	1.23	0.05
	24.53	22946.29				
	24.60	21732.73				

Table 5.2 Examples of HC copy number determination

The Absolute quantification software (Applied Biosystems) calculated HC and LC copy numbers by comparing unknown C_T values to the standard curves. Copy per cell was then calculated by dividing quantity by number of cells in sample.

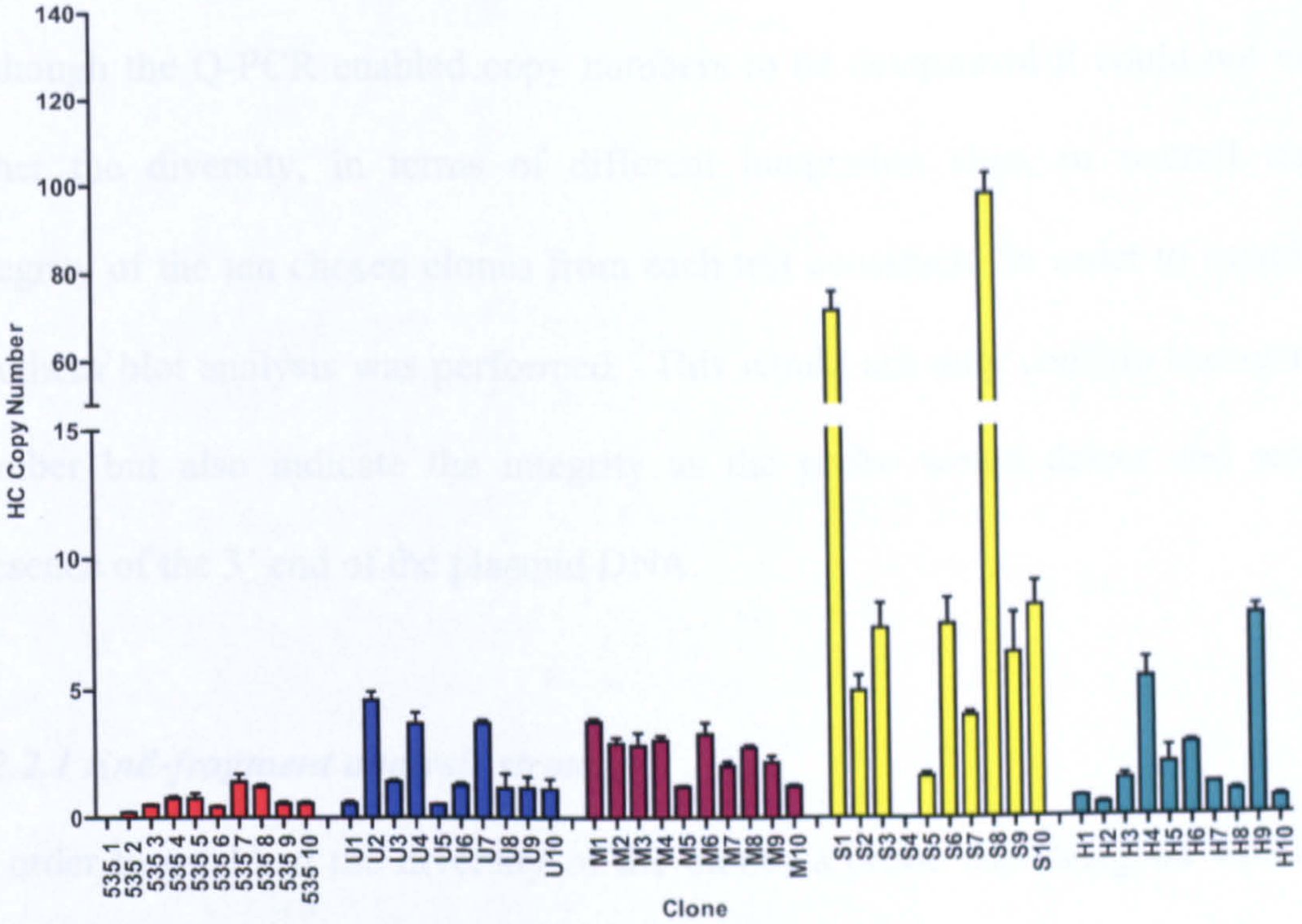
LC:

Clones	C_T	Quantity	Qty Mean	Qty SD	Copy per cell	Copy per cell SD
M1	22.30	14159.06	15162.85	1027.08	3.58	0.24
	22.21	15117.77				
	22.11	16211.73				
M2	22.87	9518.40	10983.69	1312.97	3.55	0.42
	22.61	11379.34				
	22.53	12053.35				
M3	20.56	48299.36	49574.61	1745.02	2.37	0.08
	20.55	48861.17				
	20.47	51563.32				
M4	20.01	71576.77	68879.60	7172.78	4.64	0.48
	19.95	74312.81				
	20.24	60749.22				
M5	21.50	25027.83	27035.11	2708.69	1.63	0.16
	21.44	25961.46				
	21.23	30116.04				
M6	20.19	62800.09	59396.88	2960.39	3.25	0.16
	20.32	57416.92				
	20.30	57973.61				
M7	22.40	13267.39	12668.10	864.45	1.76	0.12
	22.58	11677.14				
	22.42	13059.78				
M8	22.20	15193.28	16706.19	2025.08	3.04	0.37
	22.14	15918.53				
	21.89	19006.75				
M9	20.76	42094.71	44169.42	4202.45	2.37	0.23
	20.54	49005.75				
	20.78	41407.80				
M10	20.54	48928.98	51480.81	10264.91	2.78	0.55
	20.19	62780.93				
	20.74	42732.55				

Table 5.3 Examples of LC copy number determination
The Absolute quantification software (Applied Biosystems) calculated HC and LC copy numbers by comparing unknown C_T values to the standard curves. Copy per cell was then calculated by dividing quantity by number of cells in sample.

HC and LC gene copy numbers for each clone were calculated and are shown in Figure 5.3. Irrespective of whether chromatin modifying elements were present the copy number for the majority of clones was low with numbers below 10 for both the HC and LC genes. In two of the clones, 535 1 and S4 no HC or LC DNA was detected which indicated that the HC and LC DNA had been lost from the genome. In contrast, the copy numbers for two clones, S1 and S8, were high, with over 70 copies per cell of HC and LC being detected in these clones. Approximately equal copies of HC and LC copy numbers were detected. This was expected as the HC and LC DNA was present in the same plasmid and after integration the DNA should be co-localised in the genome.

A.



B.

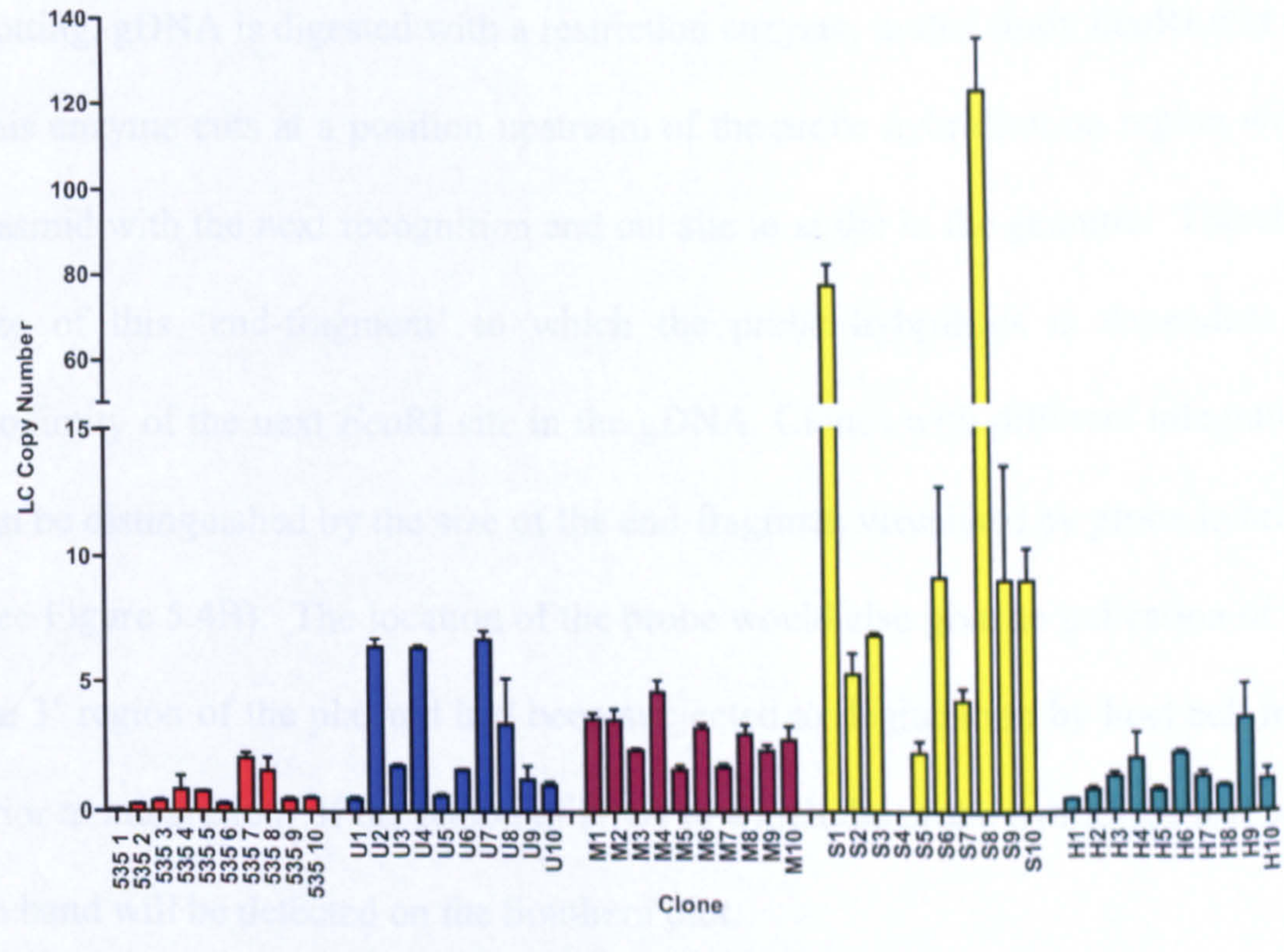


Figure 5.3 HC and LC gene copy number determination by Q-PCR
100ng gDNA was amplified with probes to the heavy and light chain. GAPDH was used as an internal control. A = HC copy number. B = LC copy number. Clones derived from vectors containing the different elements and depicted in different colours. Red = Ab535, Blue = 1.5kb A2UCOE, Purple = MAR X_S29, Yellow = STAR 7, Green = cHS4 tandem. Each data point represents mean +/- SD (n=3) of samples assayed in triplicate wells. Data shown is from one experiment.

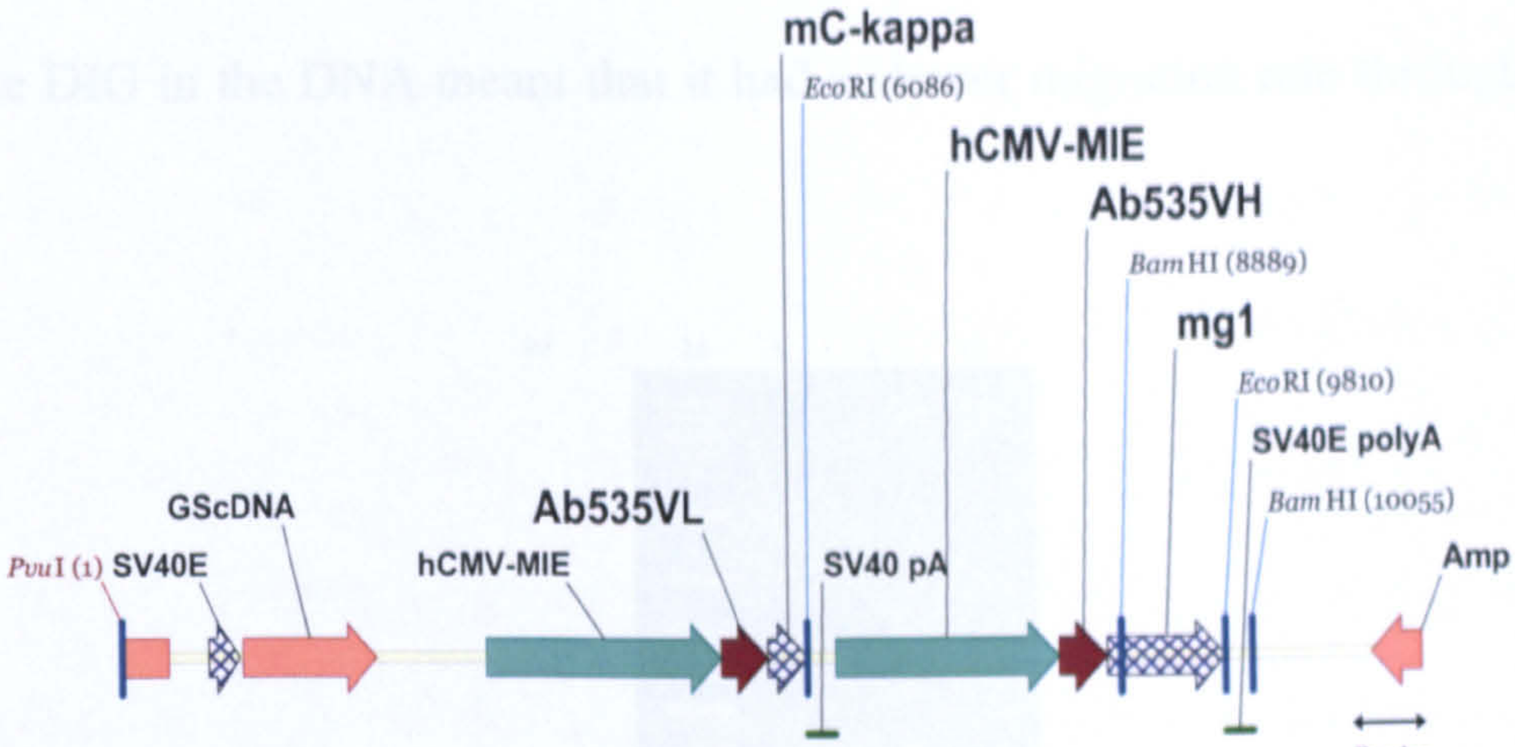
5.2.2 Confirmation of plasmid copy numbers and clonal diversity

Although the Q-PCR enabled copy numbers to be determined it could not establish either the diversity, in terms of different integration sites, or overall transgene integrity of the ten chosen clones from each test construct. In order to establish this Southern blot analysis was performed. This would not only confirm transgene copy number but also indicate the integrity as the probe would detect and assess the presence of the 3' end of the plasmid DNA.

5.2.2.1 End-fragment analysis strategy

In order to establish the diversity of the clones a probe was designed which would hybridise to the 3' end of the plasmid DNA (see Figure 5.4A). Prior to Southern blotting, gDNA is digested with a restriction enzyme, in this study *Eco*RI was chosen. This enzyme cuts at a position upstream of the probe hybridisation region within the plasmid with the next recognition and cut site in at the in the genome. Therefore, the size of this 'end-fragment' to which the probe hybridises is dependent on the proximity of the next *Eco*RI site in the gDNA. Clones with different integration sites can be distinguished by the size of the end-fragment visualised by probe hybridisation (see Figure 5.4B). The location of the probe would also give an indication of whether the 3' region of the plasmid had been subjected to degradation by host cell nucleases prior to integration. If the region of DNA to which the probe hybridises had been lost, no band will be detected on the Southern blot.

A.



B.

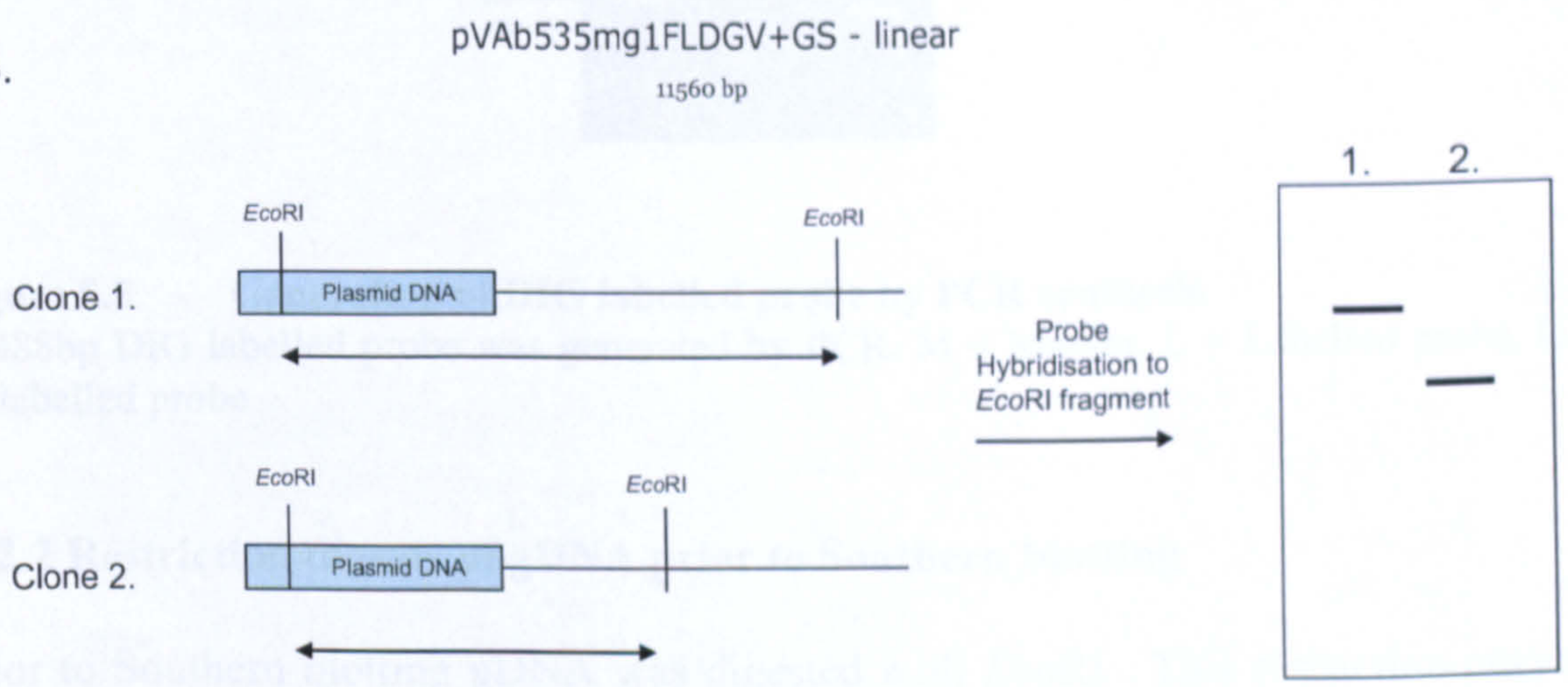


Figure 5.4 Location of probe and end fragment analysis strategy
A. Map of expression plasmid linearised with *Pvu*I. Location of probe for end fragment analysis indicated by double headed arrow.
B. Illustration of the different sized fragments generated by different integration sites. The size of fragment generated is dependent on location of *Eco*RI site in the genome.

5.2.2.2 Probe generation using PCR

The DIG labelled probe was generated using a PCR approach. In this method, DIG labelled dUTP was incorporated into the PCR product as it was generated. Oligonucleotides (see Appendix 2) were designed to generate a 488bp fragment that would hybridise at the 3' end of the plasmid as indicated in Figure 5.4A. The PCR product was generated (as described in Section 2.3.3.1) and to confirm labelling a fraction of the reaction was analysed by agarose gel electrophoresis. Figure 5.5 shows

the migration differences of the unlabelled and labelled PCR products. The presence of the DIG in the DNA meant that it had a slower migration rate through the agarose gel.

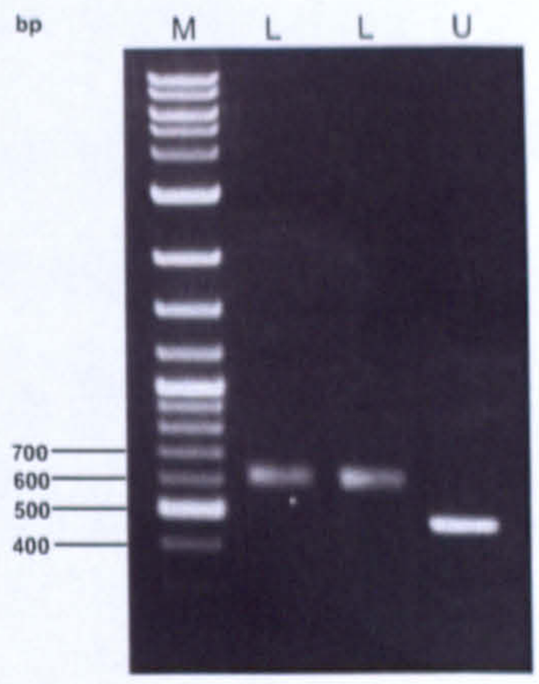


Figure 5.5 Generation of DIG labelled probe by PCR synthesis
A 488bp DIG labelled probe was generated by PCR. M = Marker, L = Labelled probe, U = Unlabelled probe

5..2.2 Restriction digests of gDNA prior to Southern blotting

Prior to Southern blotting gDNA was digested with *Eco*RI. This restriction enzyme was considered suitable based on information known of the murine genome. *Eco*RI should on average digest *Mus Musculus* DNA once every 4kb (http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/band_size_by_cleavage.asp). The Chinese Hamster genomic sequence is not in the public domain. However, given that Chinese Hamster DNA shares approximately 60% homology with the murine genome (Wlaschin et al, 2005), it was assumed that *Eco*RI would digest CHO cell gDNA with a similar frequency which would allow the resultant restricted fragment size to be of an appropriate size to be detected on the Southern blot. The CHO cell gDNA was purified using the Wizard DNA purification Kit (Promega) as described in Section 2.2.11 and agarose gel electrophoresis was performed to confirm the extracted gDNA was of a high molecular weight (Figure 5.6A). Restriction enzyme digests with *Eco*RI were set up and incubated for 16 hours

to ensure complete digestion. To confirm complete digestion, a control reaction was conducted in which bacteriophage Lambda DNA was spiked into CHO-K1 gDNA. Lambda DNA would give a known pattern of fragments upon complete digestion implying the digest reactions had been successful. A fraction of the digested DNA was visualised using agarose gel electrophoresis and a representative gel is shown in Figure 5.6B. This shows the expected smear of gDNA that would be generated due to the gDNA being fragmented into various sizes upon complete digestion. The correct pattern of digestion can also be seen for the Lambda DNA indicating a successful restriction enzyme digest.

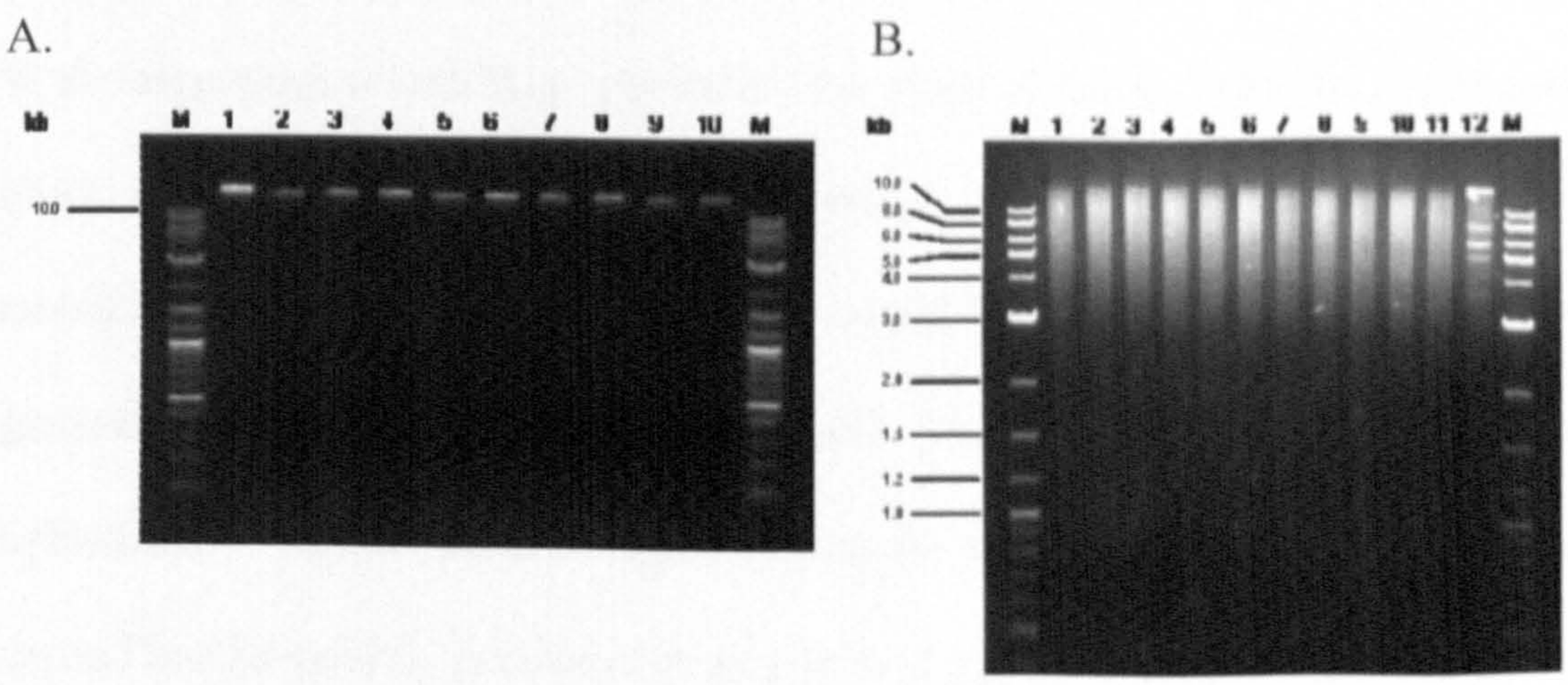
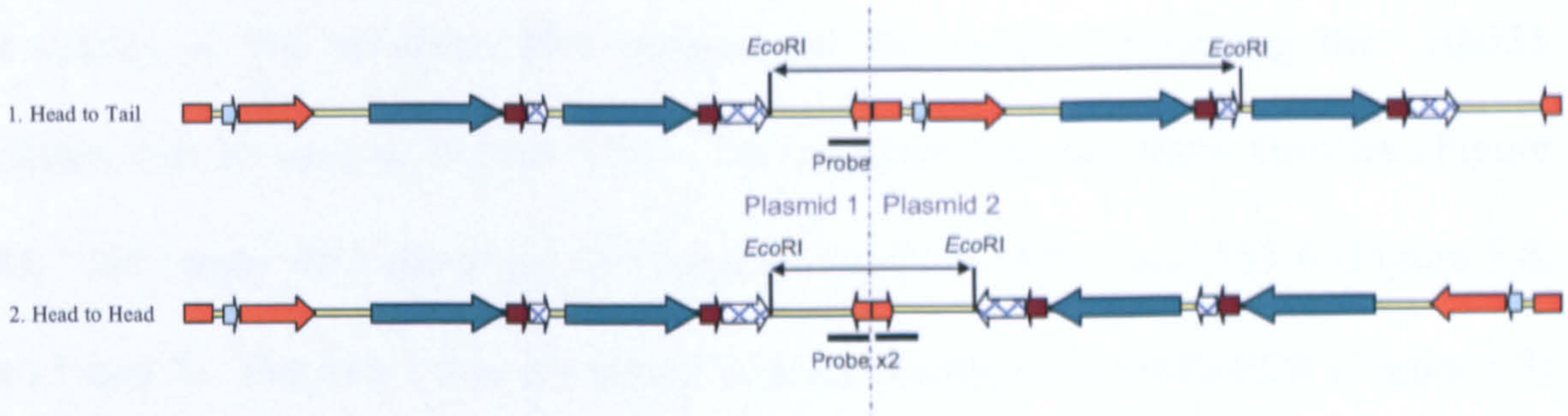


Figure 5.6 Purified and digested gDNA samples
A. 250ng of purified gDNA. Lanes 1-10 = gDNA samples. B. gDNA digested with *Eco*RI for 16 hours at 37°C. 250ng of digested gDNA was analysed by gel electrophoresis. M = Marker, Lanes 1-11 = gDNA, Lane 12 = gDNA + Lambda DNA. Expected fragments for lambda DNA = 24.7kb, 7.4kb, 5.8kb, 5.6kb, 4.9kb,

5.2.3 Southern blot hybridisation analysis

After transfection linearised DNA enters the nucleus where it can integrate into the genome at any location. In the majority of cases, if multiple copies of plasmid are integrated into the genome they insert as tandem arrays presumably due to the ligation of linearised plasmid molecules in the nucleus (Chen and Chasin, 1998). Due to the Southern blotting strategy being utilised, different integration patterns could be determined. If a single copy of the plasmid integrates into the genome, only one end-fragment band would be expected. If more than one copy of the plasmid integrates into the genome the plasmids are likely to be arranged in tandem arrays. Two possible arrangements are as a head-to-tail, or head-to-head array (see Figure 5.7). With integration of multiple copies of the plasmid there could be combinations of these arrays or other possible arrangements. With the head-to-tail tandem array example shown below, two fragments would be generated, one band would be generated from the internal *Eco*RI-*Eco*RI fragment where the intensity of the hybridisation signal would be dependent on the number of plasmids integrated and a second band would be generated from a hybrid plasmid/gDNA fragment. In the-head to-head example only the internal fragment would be generated, however the signal would be more intense as two probes would hybridise.

A.



B.

Cell line	Head-to-Tail	Head-to-Head
Ab535	7.8	3.5
1.5kb A2UCOE	10.9	6.6
MAR X_S29	6.3	3.5
STAR 7	12.1	11.4
cHS4 tandem	8.7	4.5

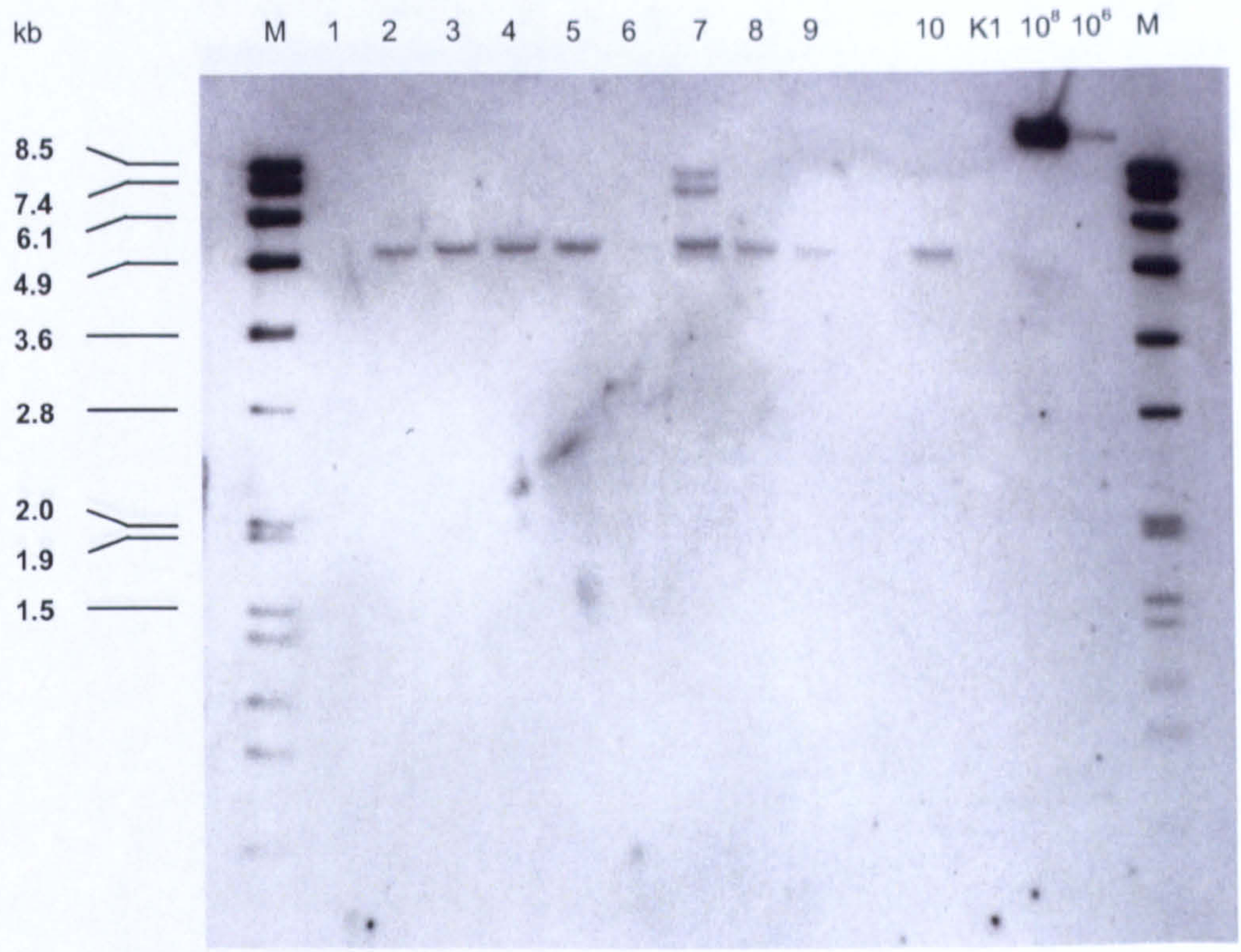
Figure 5.7 Examples of the potential integrated tandem arrays of plasmids and theoretical band sizes for internal plasmid generated fragments
A. Illustrations showing two possible tandem array formations, Head-to-Tail and Head-to-Head. B. The expected internal fragment size if tandem arrays had been formed.

Southern blotting was performed as described in Sections 2.3.3.3 to 2.3.3.7. The results for the different cell lines are shown in Figure 5.8 to Figure 5.12. In all Southern blots no bands were observed in the CHO-K1 gDNA. This demonstrated that the probe was specific for the plasmid DNA. In general, there was a lack of diversity in the clonal populations. None of the five cell lines contained ten individual clones with different integration sites. A number of clones shared the same banding patterns resulting in three to five families of clones in each cell line. The clones were isolated from pooled stable cell lines and therefore the clones which shared the same banding patterns were likely to be daughter progeny from one original tranfected cell.

5.2.3.1 Southern blot analysis for Ab535 control clones

The results of the Southern blot analysis for the clones harbouring the Ab535 construct can be seen in Figure 5.8A. These clones fell into three families (Figure 5.8B). No bands were observed for clones in family A, 535 1 and 535 6 (Figure 5.8, lanes 1 and 6). For 535 1 this correlated with the results from the Q-PCR (Figure 5.3) which also suggested no DNA was present. In contrast, the Q-PCR results for 535 6 indicated that HC and LC plasmid DNA was present and therefore the lack of band in the Southern blot may have been due to degradation of the 3' end of the plasmid DNA at or prior to the time of integration. A single band 5kb in size was observed in clones in family B: 535 2, 535 3, 535 4, 535 5, 535 8, 535, 9, and 535 10 suggesting that only one copy of the plasmid had integrated into the genome at the same site (Figure 5.8, lanes 2, 3, 4, 5, 8 and 9). Clone 535 7 showed a unique banding pattern with 3 fragments being observed. In addition to a band at approximately 5kb two further bands were present which were larger in size (approximately 8.0kb and 7.4 kb). The largest band is consistent with the size of band generated by a head to tail tandem array (Figure 5.7). However, the presence of a third band suggests that several copies of the plasmid have integrated in a different arrangement or there has been integration of another plasmid at a different location.

A.

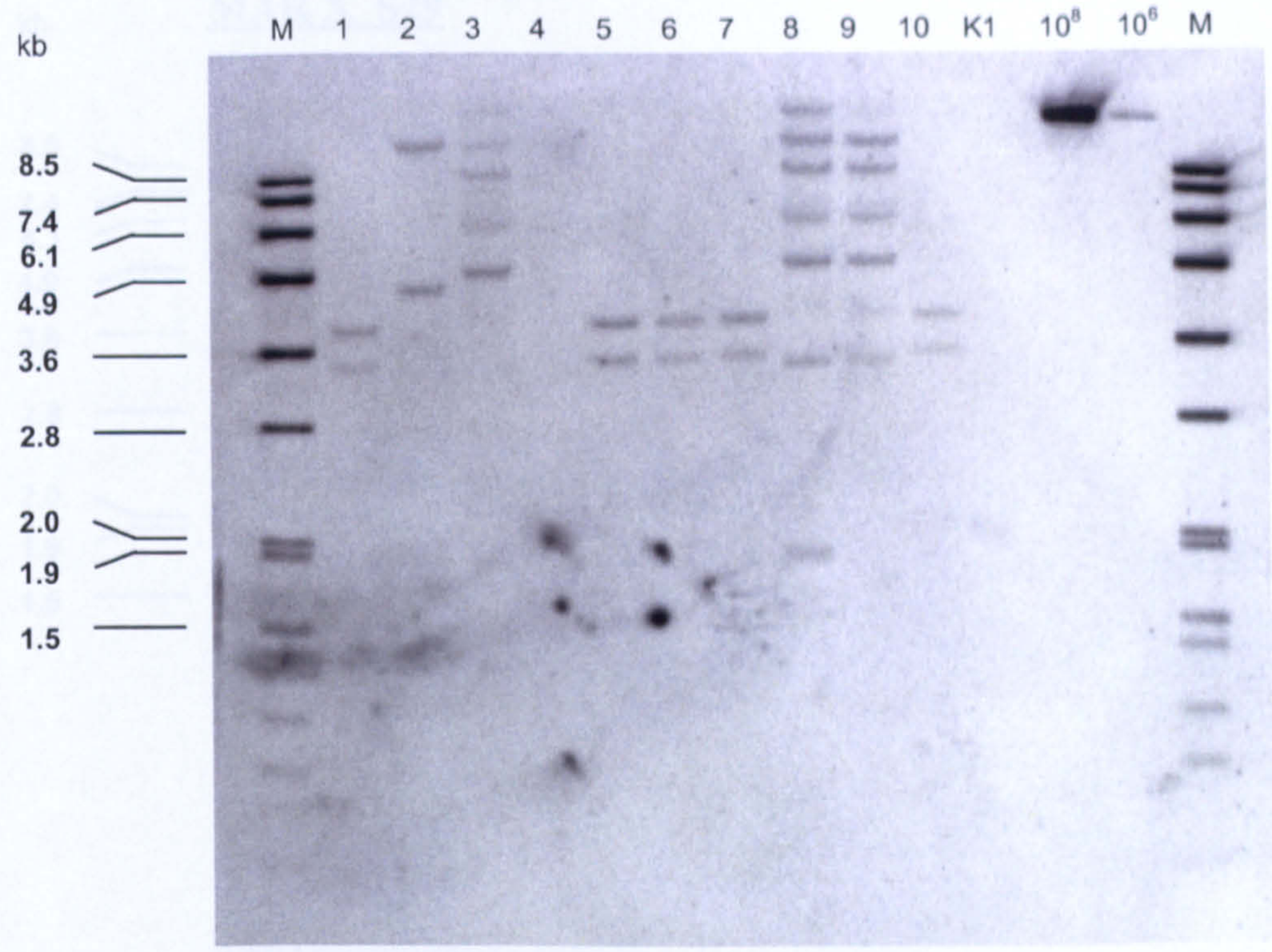


B.

Family	Clones	number of bands	Approximate band size (kb)
A	535 1, 535 6	0	
B	535 2, 535 3, 535 4, 535 5, 535 8, 535, 9, 535 10	1	5
C	535 7	3	8.0, 7.4, 5

Figure 5.8 Southern blot analysis of Ab535 clones
A. Digested DNA was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^8 = 1 x 10^8 plasmid copies. 10^6 = 1 x 10^6 plasmid copies. 1- 10 = Ab535 clones 535 1 to 535 10.
B. Families of clones as indicated by banding patterns on Southern blot

A.

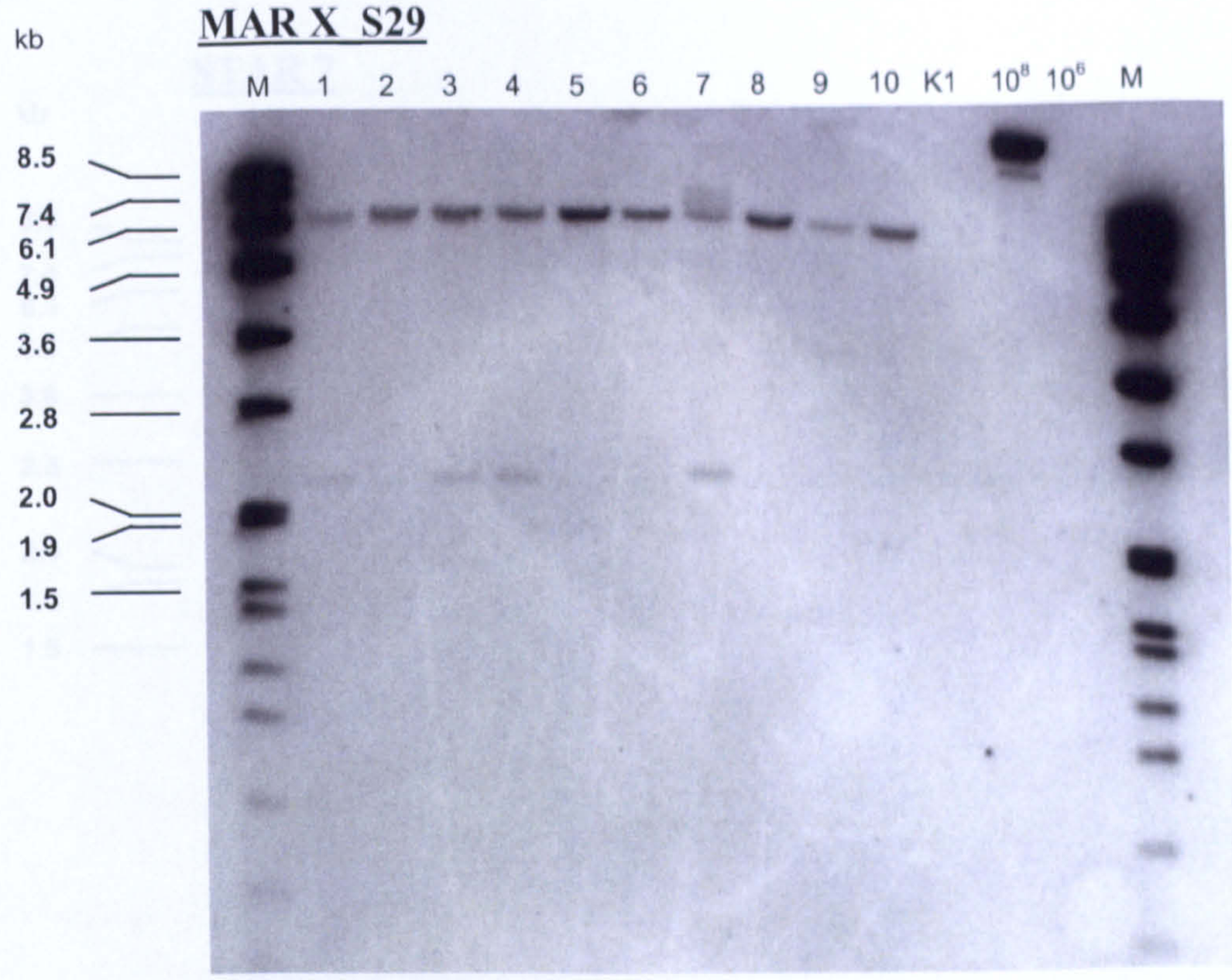


B.

Family	Clones	number of bands	Approximate band size (kb)
A	U1, U5, U6, U7, U10	2	3.4, 3.8
B	U2	2	>8.5, 4.8
C	U4	0	
D	U3, U8, U9	5 to 7	>8.5, >8.5, 8.5, 6.1, 5, 3.2, 1.9

Figure 5.9 Southern blot analysis of 1.5kb A2UCOE clones
A. Digested DNA was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^8 = 1 x 10^8 plasmid copies. 10^6 = 1 x 10^6 plasmid copies. 1- 10 = 1.5kb UCOE clones U1 to U10.
B. Families of clones as indicated by banding patterns on Southern blot

A.

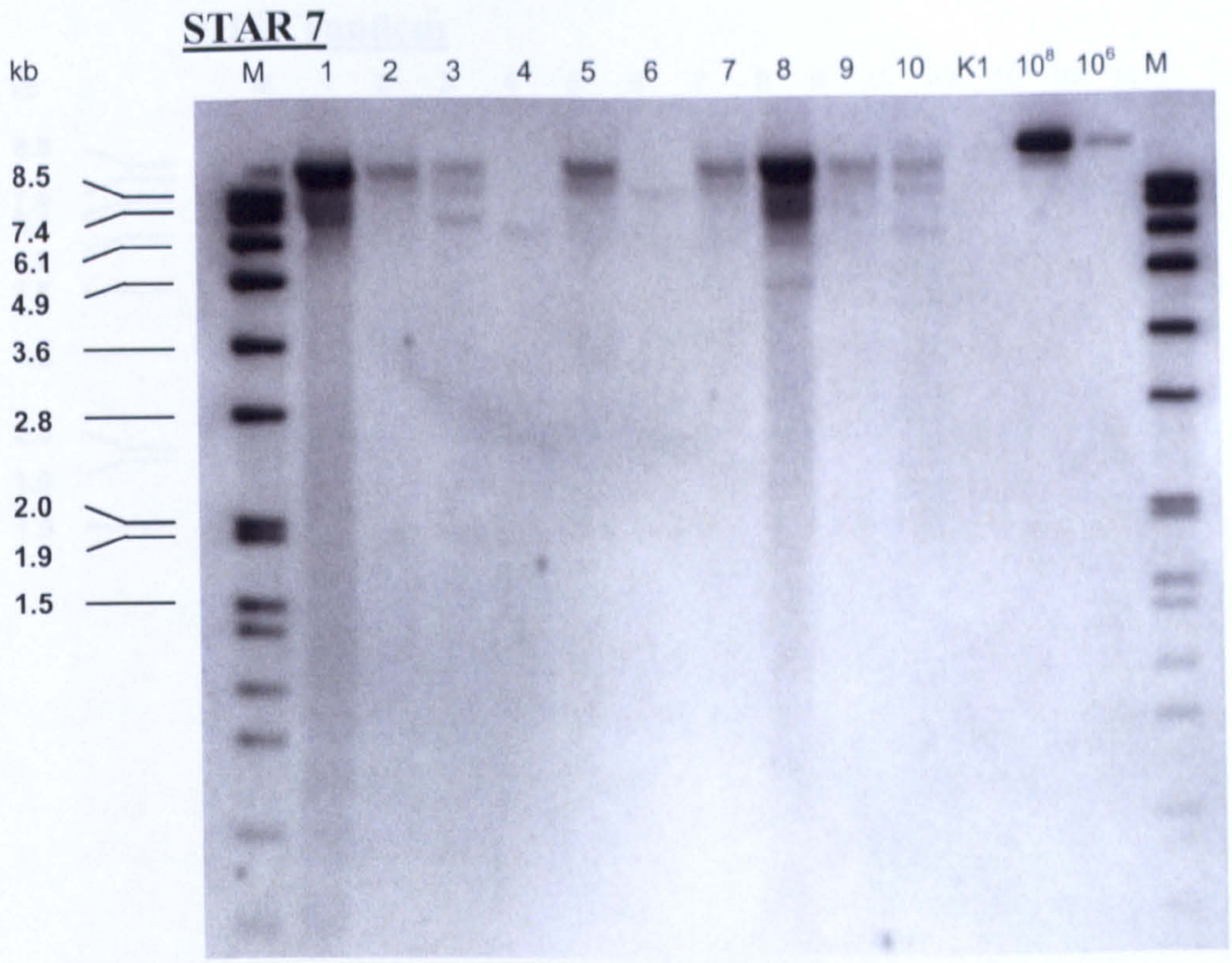


B.

Family	Clones	number of bands	Approximate band size (kb)
A	M1, M3, M4	2	6.1, 2
B	M2, M5, M6, M8, M9, M10	1	6.1
C	M7	>2	diffuse band >6.2, 6.1, 2

Figure 5.10 Southern blot analysis of MAR X_S29 clones
A. Digested DNA was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^8 = 1 x 10^8 plasmid copies. 10^6 = 1 x 10^6 plasmid copies. 1- 10 = MAR X_S29 clones M1 to M10.
B. Families of clones as indicated by banding patterns on Southern blot

A.

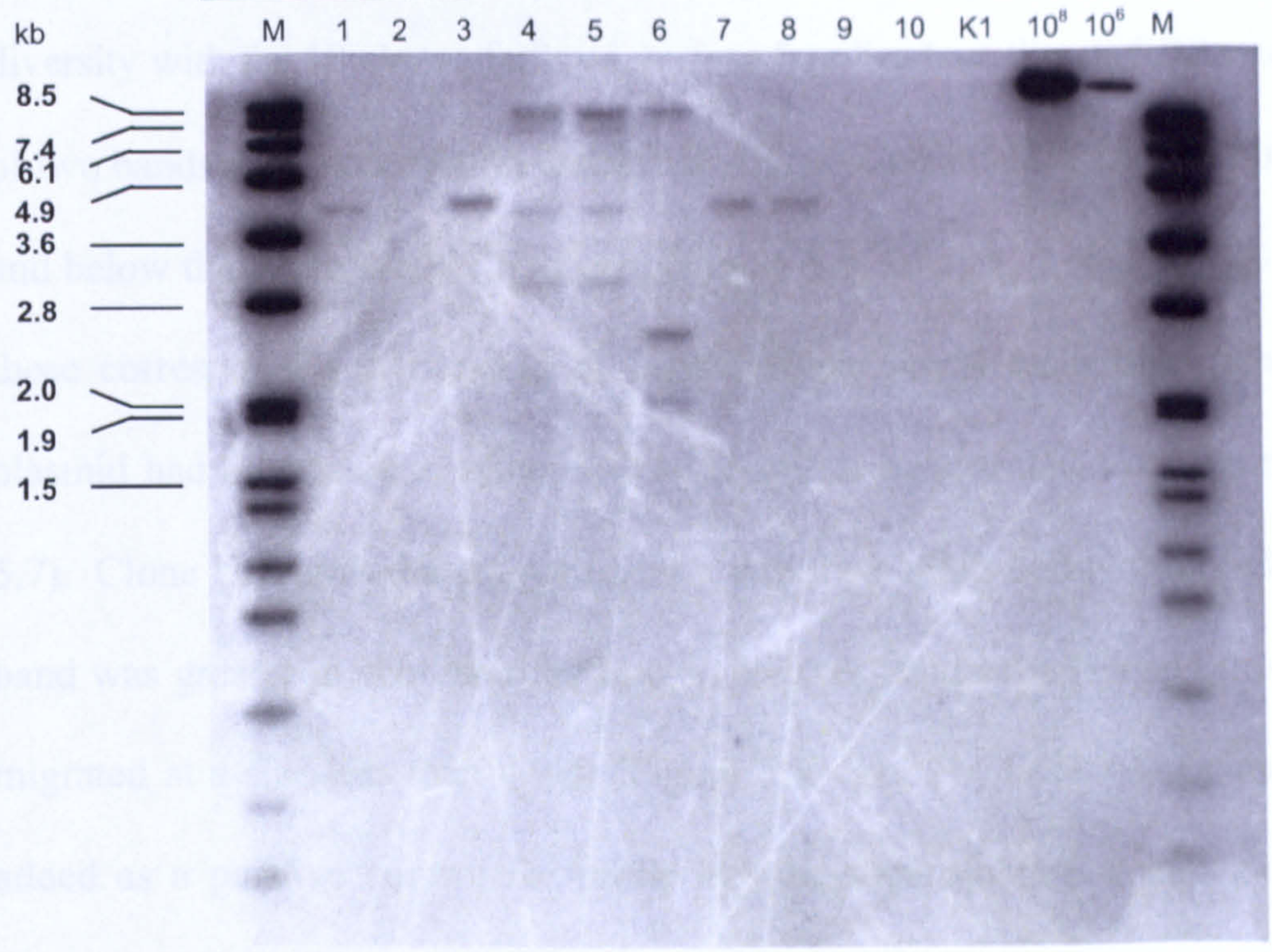


B.

Family	Clones	number of bands	Approximate band size (kb)
A	S1, S8	2 to 3	>8.5, 7.4, <4.9
B	S2, S5, S7, S9	2	>8.5, >8.5
C	S3, S10	3	>8.5, 8.5, 7
D	S4	1	6.1
E	S6	1	8.5

Figure 5.11 Southern blot analysis of STAR 7 clones
A. Digested DNA was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^8 = 1 x 10^8 plasmid copies. 10^6 = 1 x 10^6 plasmid copies. 1- 10 = STAR 7 clones S1 to S10.
B. Families of clones as indicated by banding patterns on Southern blot

A.



B.

Family	Clones	number of bands	Approximate band size (kb)
A	H1, H3, H7, H8	1	4.3
B	H2, H9, H10	0	
C	H4, H5	3	8.5, 3.6, 2.8
D	H6	3	8.5, 2.5, 1.9

Figure 5.12 Southern blot analysis of cHS4 tandem clones
A. Digested DNA was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^8 = 1 x 10^8 plasmid copies. 10^6 = 1 x 10^6 plasmid copies. 1- 10 = HS4 tandem clones H1 to H10.
B. Families of clones as indicated by banding patterns on Southern blot

5.2.3.2 Southern blot analysis for 1.5kb A2UCOE clones

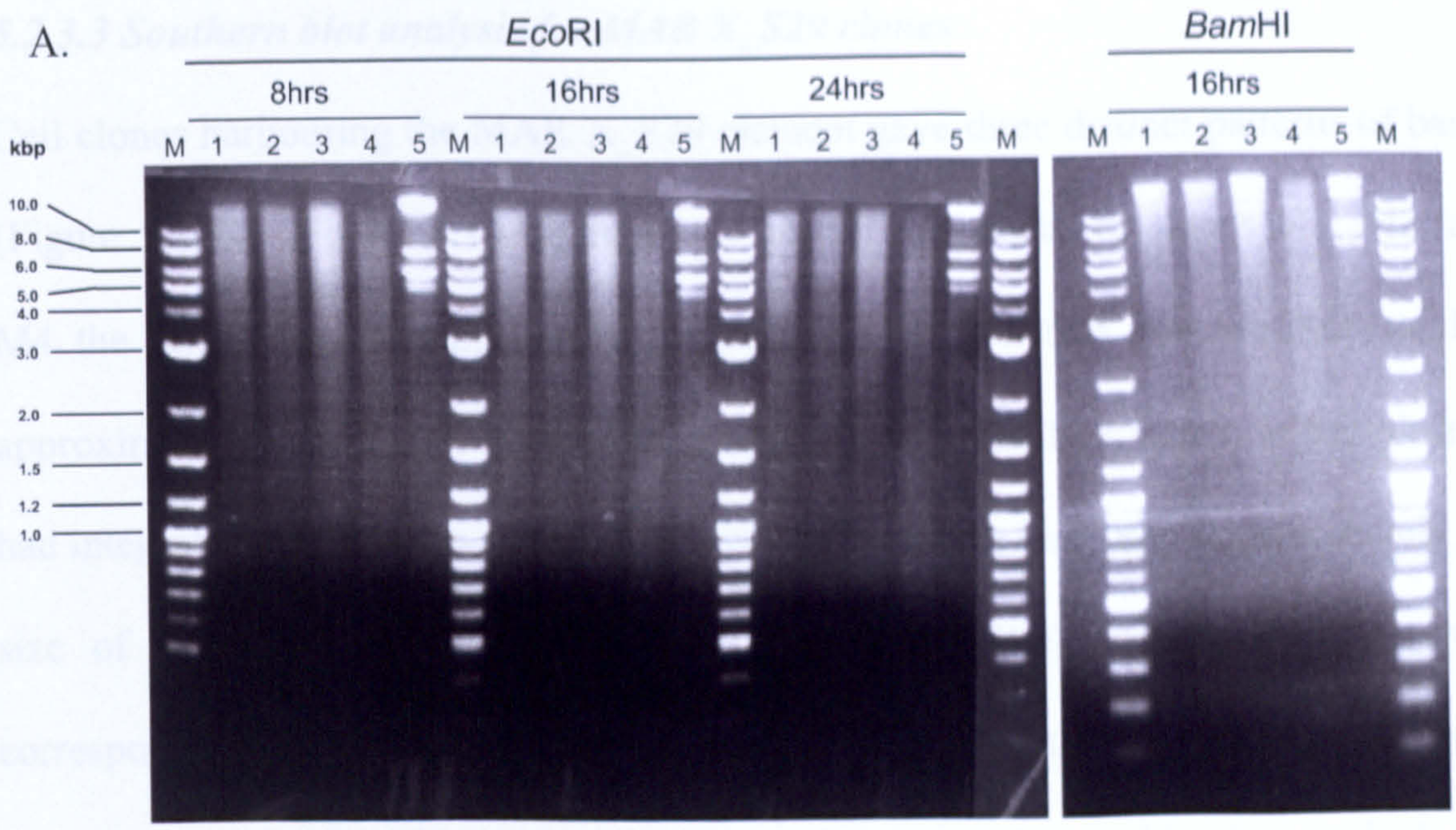
The results of the Southern blot analysis for the 1.5kb A2UCOE clones showed more diversity with the 10 clones falling into four families (see Figure 5.9A and B). A total of two bands were observed in clones U1, U5, U6, U7 and U10 which migrated above and below the 3.6kb marker fragment (Figure 5.9, lanes 1, 5, 6, 7 and 10). Neither of these corresponded to the size of bands which would have been generated if the plasmid had integrated in either a head-to-tail or head-to-head configuration (Figure 5.7). Clone U2 showed a unique pattern, with two bands being observed. The largest band was greater in size than the biggest marker band of 8.5kb and the second band migrated at a size less than 4.9kb (Figure 5.9, lane 2). Linearised plasmid was also added as a positive control for probe hybridisation and had a molecular weight of 16.2kb. The largest band was below this and could therefore correspond to a 10.9kb band generated if the plasmids had integrated as a head-to-tail tandem array. No bands were observed in U4 (Figure 5.9, lane 4). The results from the Q-PCR of this clones suggested HC and LC DNA had integrated into the genome (Figure 5.3A and B) and therefore the lack of band in the Southern blot was probably due to the 3' end of the linearised plasmid being digested by host cell nucleases resulting in the removal of the region where the probe hybridised. The remaining three clones, U3, U8 and U9 showed a more complex pattern. A total of seven bands were observed in U8 ranging from approximately 16kb to 3.6kb in size (Figure 5.9, lanes 3, 8 and 9). Although fewer bands were observed in U3 and U9, the band sizes were the same as U8 which suggested that all three clones were daughter progeny from one original transfected cell. If the plasmids had integrated as tandem arrays a more intense band at either 10.9kb or 6.6kb would have been observed but this was not the case. A possible cause may have been partial digestion of the gDNA although restriction enzyme

digested control lambda DNA showed a complete digestion pattern suggesting incomplete digestion was not the cause.

To investigate whether the number of bands were due to under or over digestion of the gDNA the digest reaction time was altered. In addition, an alternative restriction enzyme *Bam*HI was used, which had a recognition site in close proximity to the *Eco*RI site in the plasmid (see Figure 5.4A.)

Clone U8 was chosen as a representative example of the clones generating the multiple banding pattern for further investigation. Clones U2 and U7 with different hybridisation signals were also included. Restriction digests were performed where the incubation time was reduced to 8 hours or increased to 24 hours. The results of the gDNA digests after resolution by agarose gel electrophoresis are shown in Figure 5.13A. The *Eco*RI restriction enzyme digest pattern looked similar for all three incubation times and the correct sized fragments for the Lambda DNA were observed (Figure 5.13, lanes 5 in *Eco*RI digests). The digestion of gDNA with *Bam*HI resulted in larger gDNA fragments being generated as indicated by the increase in high molecular weight bands at the top of the 'smear' of DNA. The Lambda DNA was digested into the expected fragment sizes, which would suggest the digest had been successful. The frequency of *Bam*HI sites in the *Mus Musculus* genome is the same as *Eco*RI sites (on average, once every 4kb), however, in light of the results of the restriction digests performed it would appear that the frequency of *Bam*HI sites is lower in the CHO genome.

The digested gDNA was analysed by Southern blot hybridisation and the results are shown in Figure 5.13.B. Irrespective of the incubation time, the banding patterns for the three clones were similar to the results obtained previously (Figure 5.9A). The digestion of gDNA for 8, 16 or 24 hours all resulted in the multiple hybridisation signals in clone U8 (Figure 5.13, lanes 3 in *Eco*RI digests). This indicated that the bands were not due to under or over digestion of the DNA. The multiple bands were also present when the gDNA had been digested with *Bam*HI (Figure 5.13, lane 3 *Bam*HI digest). This suggested that the multiple bands generated in clone U8 were not an artefact of the Southern blot but could be due to multiple copies of the plasmid integrated at different locations in the host cell genome.



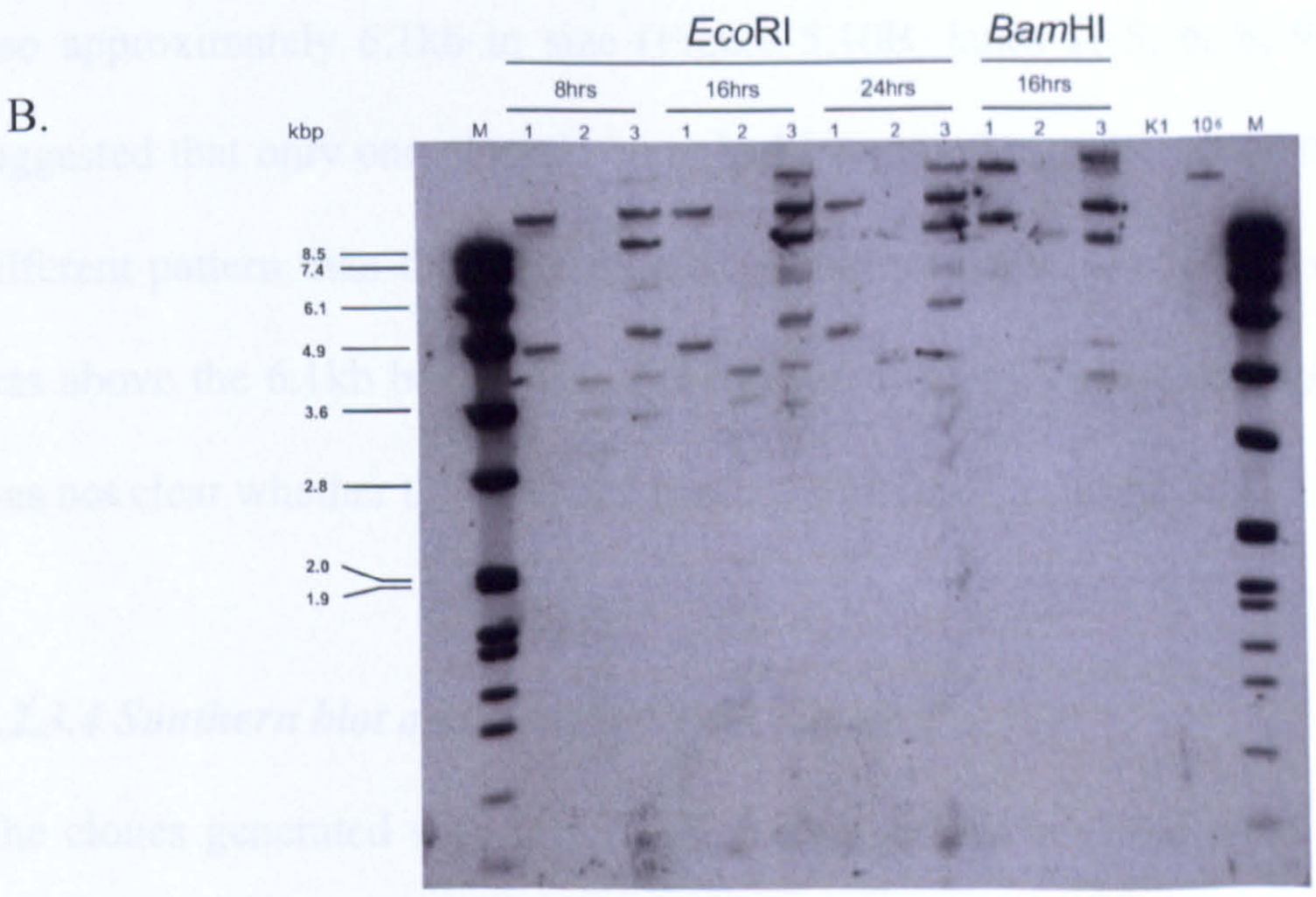


Figure 5.13 Altering incubation time and restriction enzyme for gDNA digestion
A. 250ng of gDNA digested with *Eco*RI or *Bam*HI for 8, 16 and 24 hours and visualised by agarose gel electrophoresis 1 = U2, 2 = U7, 3 = U8, 4 = CHO-K1, 5 = CHO-K1 + Lambda DNA. Expected fragments for lambda DNA = 24.7kp, 7.4kb, 5.8kb, 5.6kb, 4.9kb. Expected fragments for lambda DNA digested with *Bam*HI = 16.8kp, 12.2kb, 7.2kb, 6.5kb, 5.6kb.
B. Southern blot hybridisation using DIG labelled probe 1 = U2, 2 = U7, 3 = U8, K1 = CHO-K1, 10^6 = 1 x 10^6 copies of UCOE GS plasmid

5.2.3.3 Southern blot analysis for MAR X_S29 clones

Cell clones harbouring the MAR X_S29 element gave three distinct patterns of bands (Figure 5.10B). Figure 5.10A (lanes 1, 3 and 4) shows that in clones M1, M3, and M4 the larger band is approximately 6.1kb in size whereas the smallest band is approximately 2kb in size. The band sizes suggested that two copies of the plasmid had integrated as a head-to-tail tandem array as the larger band was approximately the size of the band size of the internal fragment (6.3kb) and the 2kb fragment corresponded to the vector/gDNA junctional end-fragment. In the second family of clones (M2, M5, M6, M8, M9 and M10), one band was observed in clones which was also approximately 6.1kb in size (Figure 5.10B, lanes 2, 5, 6, 8, 9 and 10). This suggested that only one plasmid copy had integrated into the genome. M7 showed a different pattern than the other clones. A further diffuse band was observed, which was above the 6.1kb band. The diffuseness of the hybridisation signal meant that it was not clear whether this was one band or several of a similar size.

5.2.3.4 Southern blot analysis for STAR 7 clones

The clones generated with the STAR 7 construct showed the most diversity, falling into five families (Figure 5.11). The signal intensity from clones S1 and S8 was high, indicating that a large number of plasmid copies had integrated into the genome in these clones (Figure 5.11, lanes 1 and 8). The size of the largest band was smaller than the plasmid DNA control (15.8kb) and may therefore correlate with the expected restriction fragment (12.1kb) generated from a head-to-tail tandem array. The Southern blot confirmed data generated from the Q-PCR (Figure 5.4) in which the HC and LC copy number estimate was over 70 for both these clones. Clones in family B, S2, S5, S7 and S9, had two bands, which were both larger than 8.5kb in size (Figure

5.11, lanes 2, 5, 7 and 9). The largest molecular weight band was the same size as the largest band in clones S1 and S8. Clones S3 and S10 shared the same banding pattern, with three bands being observed. Again, the largest band was the same size (>8.5kb) as the largest band observed in the clones in families A and B indicating that plasmids had integrated as a head-to-tail tandem array. In contrast to clones S1 and S8 the hybridisation signal intensity was lower in the clones in families B and C, which indicates that fewer plasmid copies had integrated (Figure 5.11, lanes 3 and 10). Single hybridisation signals were observed in clones S1 and S4 but these were of different sizes. This suggests that these clones contain a single plasmid copy but the integration site was different in each. The observation of a band in S4 suggesting plasmid DNA had integrated in the genome disagreed with the results of the Q-PCR (Figure 5.3) where no HC or LC DNA was detected, suggesting that degradation of the plasmid had occurred and only small regions of the plasmid remained integrated in the genome.

5.2.3.5 Southern blot analysis for HS4 tandem clones

The banding patterns for clones harbouring the cHS4 construct fell into four groups (Figure 5.12). One band was observed in clones H1, H3, H7 and H8, which was between 3.6kb and 4.9kb in size (Figure 5.12, lanes 1, 3, 7 and 8). This indicated that a single plasmid copy had integrated in these clones. No bands were observed for clones H2, H9 and H10 (Figure 5.12, lanes 2, 9 and 10). In contrast to the Southern blot analysis, the Q-PCR results (Figure 5.3) indicated that HC and LC DNA was present. The lack of bands observed in the Southern blot suggested that partial degradation of the plasmid had occurred in these clones and the 3' region of the plasmid where the probe hybridised was absent. In clones H4 and H5 three bands

were observed, which were approximately 8.5kb, 3.6kb and 2.8kb in size (Figure 5.12, lanes 4 and 5). The largest band corresponded to the expected fragment size generated by a head-to-tail tandem array (8.7kb). However, only two bands would be expected if the plasmids were integrated in a tandem array. A possible explanation for this is that there are separate integration sites for the plasmid DNA in this clone or that partial regions of the plasmid have been integrated. Clone H6 showed a unique pattern of bands with three bands being seen, the largest being approximately 8.5kb and the smaller two bands were approximately 2.5kb and 1.9kb in size (Figure 5.12, lane 6).

5.2.4 Mapping integration sites by FISH

The multiple bands generated in the Southern blot hybridisations for clone U8 (Figure 5.9, lane 8) could have been a result of multiple integration events at different locations in the genome. To determine whether this was the case, metaphase FISH was performed on three 1.5kb A2UCOE clones (U2, U7 and U8). FISH allows the site of integration within the chromsomes to be visualised.

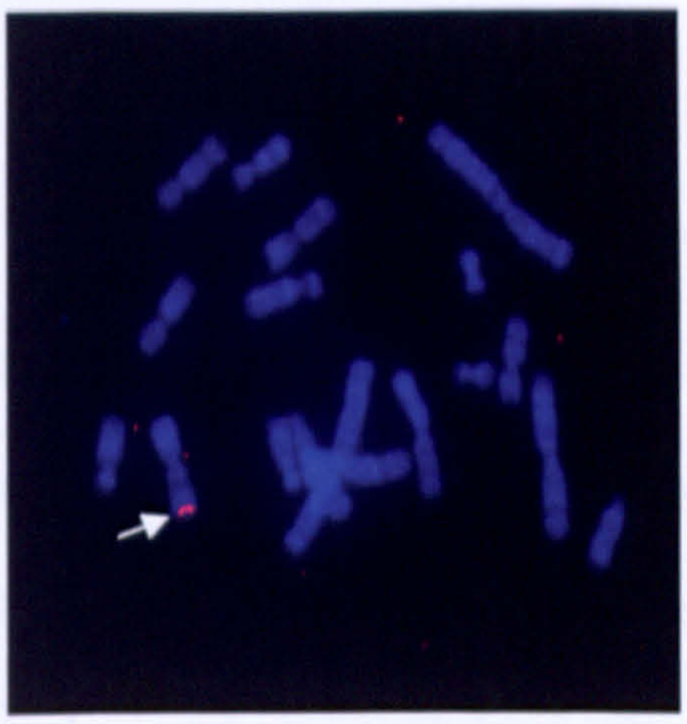
Metaphase FISH was performed as described in Section 2.6.1 with a biotinylated probe prepared from the 1.5kb A2UCOEGS plasmid DNA. The probe consisted of the complete plasmid, which had been labelled and digested by DNaseI as described in Section 2.6.1.1. At least 10 metaphase spreads were analysed for each clone and untransfected CHO-K1 cells with representative images shown in Figure 5.14. The lack of signal seen in CHO-K1 chromosomes meant that the probe was specific for plasmid DNA (Figure 5.14D). For all three clones only one hybridisation signal was

observed suggesting that for each clone there was only one integration site (Figure 5.14A-C).

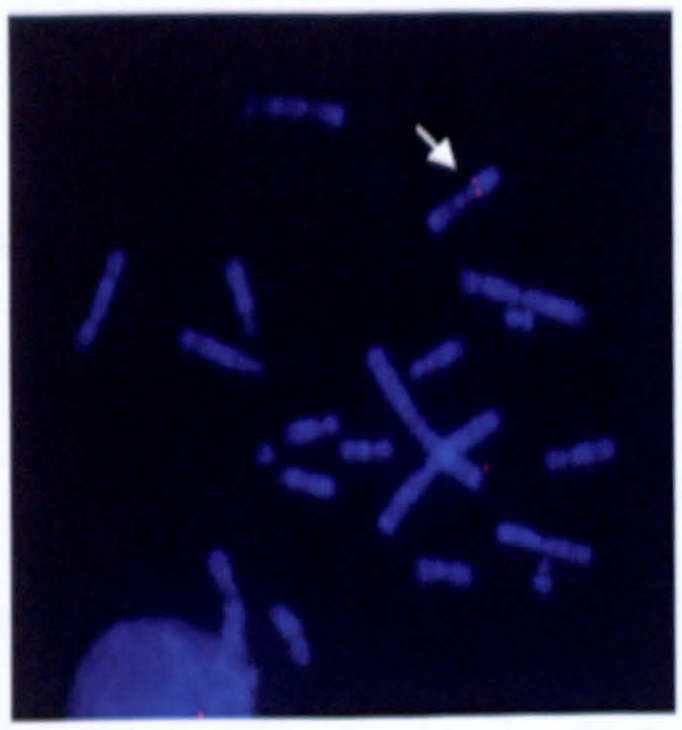
In the light of these results obtained by FISH, the multiple hybridisation signals in the Southern blot analysis of U8 may have been due to integration of multiple copies of the plasmid, which were not fully intact. If deletions or rearrangements of the DNA had occurred before integration different sized fragments would be generated upon restriction digestion and visualisation by Southern blot analysis.

In clones U2 (Figure 5.14A) and U8 (Figure 5.14C) hybridisation signals appear in telomeric and centromeric regions respectively and therefore transcriptionally repressive heterochromatin environments. Normally, little or no gene expression would be expected from such events. However, this is not the case with these clones with both showing efficient antibody production (Figure 6.1B). The observed stable transgene expression in the U2 and U8 clones is in accord with previous observations, which showed that the A2UCOE can allow expression even from within heterochromatin integration sites (Antoniou et al, 2003).

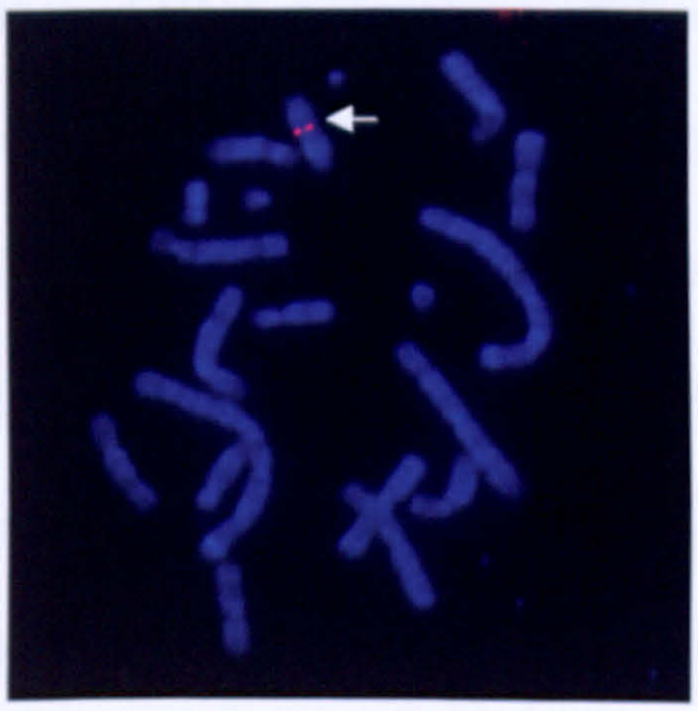
A. U2



B. U7



C.U8



D. CHO-K1

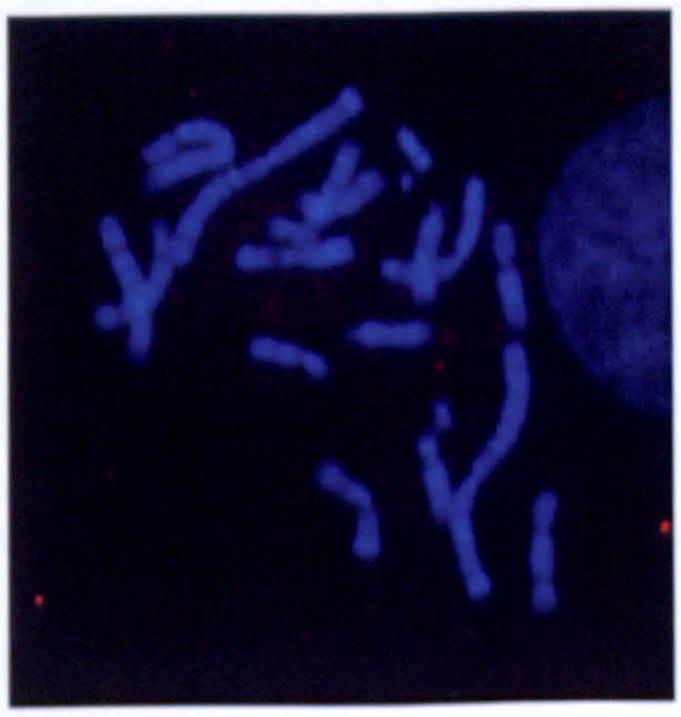


Figure 5.14 Mapping of integration sites by Fluorescence In Situ Hybridisation
Biotinylated probes were generated using the 1.5kb A2UCOE plasmid and hybridised to metaphase spreads for the three 1.5kb A2UCOE clones. Probes were detected with streptavidin conjugated to Cy3 and integration sites are indicated by pink dots (highlighted by white arrow). Chromosomes were counterstained with DAPI
A= U2. B= U8. C= U9. D= CHO-K1.

5.3 Discussion

This chapter has described the work undertaken to characterise on a molecular genetic level ten cell clones for each of the five test constructs. Q-PCR was performed to quantify the HC and LC gene copy number. A 1:1 ratio for the HC and LC genes was observed and for the majority of clones the copy number was low. The low copy number was expected due to the transfection method utilised, namely electroporation. This method genertates transient pores within with plasma cell membrane allowing plasmid DNA molecules to diffuse into the cell, some of which will be integrated into the genome. The exception was the two STAR 7 construct-containing clones, S1 and S8, where the copy number for the HC and LC genes was found to be very high (>70 copies). This increase in the number of copies was confirmed in the Southern blot analysis. STAR elements are purported to confer copy number-dependent expression (Kwaks *et al.*, 2003). However, the results presented here would suggest that this is not the case. If STAR 7 did confer copy number dependent expression the expression levels of the S1 and S8 clones would be expected to be approximately 70 times higher than the level from clones S4 or S6 which had a single plasmid copy integrated, which is not the case (see Section 6.2.1).

The Southern blot analysis was performed to establish clonal diversity, transgene integrity and confirm copy number estimations from the Q-PCR. The Southern blot analysis highlighted a lack of diversity among clones. None of the ten clones harbouring a given test construct showed integration events at different sites. In each set of cell clones several shared the same banding patterns indicating that a number of the clones investigated were probably daughter progeny from the same original

transfected cell. The clones analysed were isolated from the pools of stably transfected cells (see Section 4.2.5). Cells from these pools were plated into semi-solid CloneMedia after cell numbers and viability had recovered (approximately 3 weeks post transfection). The results of the Southern blot hybridisations suggests that by this time the diversity of the pools was low, with many of the cells in the population being daughter progeny rather than individual clones. In light of this result, it would appear that the time point at which clones are isolated from a pool of stably transfected cells is important for achieving a diverse population, with early plating after selection being crucial.

The hybridisation pattern for the three 1.5kb A2UCOE construct containing clones, U3, U8 and U9, gave multiple bands suggesting more than one integration site, which was unexpected. It is generally observed that if multiple transgene copies are integrated into the host cell genome, they do so at one integration site as tandem arrays (Chen and Chasin, 1998). Therefore, only two bands would be expected upon Southern blot hybridisation in the form of end-fragment analysis; one as a result of the internal plasmid fragment whose intensity would be dependent on the number of plasmid copies integrated (see Figure 5.6) and another from a plasmid/gDNA junctional end-fragment. Additional Southern blot analysis was performed to investigate whether the presence of multiple bands was a result of under or over restriction enzyme digestion. Irrespective of incubation time, multiple bands were observed as before. Multiple bands were also seen when gDNA was digested with a different restriction enzyme (*Bam*HI) prior to Southern blotting. This indicated that the multiple bands were not an artefact of the Southern blot procedure.

Although it is generally thought that plasmids co-integrate together in the genome, there have been reports where multiple integration events have been observed. In a study by Derouazi *et al.*, 2006, metaphase FISH analysis was employed to visualise plasmid integration sites with one clone found to have two integration sites on separate chromosomes. Therefore, it was not entirely unexpected that the multiple hybridisation signals may have been due to multiple integration sites. To confirm whether this was indeed the case metaphase FISH was performed. Hybridisation of a 1.5kb A2UCOE plasmid probe to metaphase spreads allowed the integration sites to be localised. In all three clones analysed harbouring the A2UCOE construct only one signal was evident, demonstrating that transgenes had integrated at a single site. Clearly, the multiple hybridisation signals on the Southern blot were not due to multiple integration sites. In two of the clones, U2 and U8, the results of the FISH suggested that the plasmid DNA had integrated into heterochromatic regions of the genome (U2 into a telomere and U8 into a centromere) from which expression would not normally be expected. However, the observed high-level antibody production seen with these clones confirms the potency of the A2UCOE to allow efficient expression even from within a transcriptionally repressive heterochromatin environment (Antoniou *et al.*, 2003). However, it should be noted that the results of the FISH need to be confirmed with double labelling FISH experiments where centromeric or telomeric regions are probed along with the transgene to definitively show co-localisation.

Within the scope of this study the exact nature of the integrated plasmid DNA could not be established. However, a possible explanation for the multiple bands may have

been due to the integration of partially degraded plasmid DNA that resulted in the different banding patterns observed.

In summary, this Chapter has described the genetic characterisation of the clones harbouring each of the five test constructs. Gene copy number and clonal diversity was established. Further analysis of these clones was subsequently undertaken, to investigate how the presence of chromatin modifying elements in the expression vector were affecting antibody expression from these clones, which is described in Chapter 6.

CHAPTER 6

Stability Of Clonal Cell Lines And Epigenetic Analysis

6.1 Introduction

For a successful manufacturing cell line, the expression of the recombinant gene must be constant over a long period of time. During large scale manufacturing, culture volumes of up to 20,000l can be used and therefore the generation number of cell lines when they reach the final culture volume is high. At large scale culture volumes it is not cost effective to use selection pressure. In addition, the presence of inhibitors such as MSX can cause toxicity to cells. Subsequent requirement of additional steps in downstream processing and its presence in the final culture reactor may prove to be problematic for regulatory approval of the final product. Therefore, it is essential for the expression level of a cell line making a therapeutic product to remain stable in the absence of selection pressure.

The aim of this part of the study was to assess whether the presence of the chromatin modifying elements within the expression vectors would influence stability of the established clonal stable cell lines and to understand whether epigenetic regulation, in the form of DNA methylation was contributing to any loss in expression.

Initially, clones were cultured in 6 well plates for a total of 120 generations in the absence of MSX in the culture medium. At certain time points throughout the culture period mRNA and antibody expression levels were determined. In order to ascertain whether any decrease in protein and mRNA levels were the result of the loss of gene copies from the genome Southern blot analysis was performed on the clones at the end of the stability study and compared with the Southern blot analysis undertaken in Chapter 5. The methylation status of a region of the hCMV-MIE promoter from both the HC and LC transcription units was investigated by bisulphite conversion and

DNA sequencing to establish whether loss of expression was accompanied by an increase in DNA methylation.

6.2 Results

6.2.1 Analysis of antibody expression levels during long-term culture in the absence of selection pressure

The clones chosen for analysis were the ten from each of the five cell lines which were genetically characterised (see Chapter 5). The stability of antibody expression was assessed by culturing cells in 6 well plates in the absence of MSX selective pressure. Cells were continuously passaged for 120 generations and antibody productivity was assessed with overgrow cultures set up at 0, 40, 80 and 120 generations (with 0 generations being the point at which MSX was removed from the culture medium, approximately four weeks after single colonies were isolated using the Clonepix^{FL}). For overgrow cultures cells were seeded at 3 x 10^5 cells/ml in a total of 3mls and cultured for 10 days. Supernatant was assayed using a mouse IgG ELISA as described in Section 2.5.1 and the results are shown in Figure 6.1.

Antibody expression was not detected in clone 535 1. This was not unexpected given the results of the Q-PCR and Southern blot hybridisations (Figure 5.3 and Figure 5.8) where no plasmid DNA was detected. Expression levels for 535 6 were very low with expression barely detectable at 0 generations. This again was expected given that no bands were detected on the Southern blot analysis and less than one copy of HC and LC DNA being detected by Q-PCR (Figure 5.3 and 5.8). At the start of this stability study similar expression levels, of approximately 5µg/ml were observed for clones

535 2, 535 3, 535 4, 535 5 and 535 9. After 40 generations the expression levels of these clones decreased to 1µg/ml or lower. In the subsequent 80 generations no further decline in expression was observed. Clones 535 7, 535 8 and 535 10 had higher expression levels at 0 generations with antibody titres above 10µg/ml. However, expression levels in clones 8 and 10 decreased in a similar way to clones 535 2, 535 3, 535 4, 535 5 and 535 9. With the exception of 535 7 these clones belonged to one family as assessed by having the same sized single band in the Southern blot analysis. In contrast to the other clones, the expression levels for clone 7 remained stable until 80 generations. Subsequently there was a decrease in expression and by 120 generations only 24% of the initial expression remained. This difference may be attributed to the fact that a different banding pattern was observed in this clone which indicated that the plasmid DNA had integrated at a different location in the genome which may have been more favourable for stable expression.

Overall, the 1.5kb A2UCOE clones had the highest expression levels with antibody titres above 30µg/ml for several clones at 0 generations. Clones U1, U7 and U10 had over 75% of the starting expression remaining at 120 generations, with 81%, 79% and 80% respectively. A further two clones, U5 and U6 also showed a similar expression profile over the 120 generations. However, there was a slightly larger decline in expression with 69% and 60% of starting expression remaining. These five clones could be grouped into one family as the same banding pattern was observed in the Southern blot analysis (Figure 5.10) and therefore were likely to be daughter progeny which would explain why these five clones shared a similar expression profile. The remaining clones saw a larger decline in expression levels over the 120 generations. However, with the exception of U3, U4, and U8 which had less than 50% of the initial

expression level remaining, the decline in expression levels was less evident than in the control Ab535 clones. From the banding pattern observed in the Southern blot hybridisation (Figure 5.10) U3, U8 and U9 all appeared to be daughter clones with multiple hybridisation signals present. Although U9 shared the same pattern of hybridisation signals as U3 and U8 in the Southern blot analysis, the decrease in expression was less pronounced with 61% of expression remaining at 120 generations compared to 44% and 28% for U3 and U8 respectively. Expression levels for U4 were approximately 20µg/ml at 0 generations. However, by 120 generations only 22% of the starting expression level remained. No bands were detected in this clone (Figure 5.9), however HC and LC DNA was detected by Q-PCR (Figure 5.4) and therefore the plasmid DNA in this clone may have been partially degraded before integration resulting in the loss of an A2UCOE from the vector.

At the start of the stability study clones M1, M3, M4 and M7 had the highest expression levels of the MAR X_S29 clones. After 40 generations there was a decrease in antibody titres detected for M1 and M7. Subsequent to this initial decrease in expression the levels remained constant for M1 during the remainder of the study. There was a further 2-fold decrease in expression in M7 by 120 generations. A decrease in expression was also observed in clones M3 and M4. However, this was more gradual with expression levels at 40 generations similar to those observed at 0 generations. At 80 generations there was a decrease in expression which continued at 120 generations with only 23% and 11% of expression remaining respectively for the two clones. The gradual decline in expression was also observed in clones M5, M6, M8, M9 and M10, however, the expression levels at the start of the study were lower than M3 and M4. Although M2 had the same banding pattern as

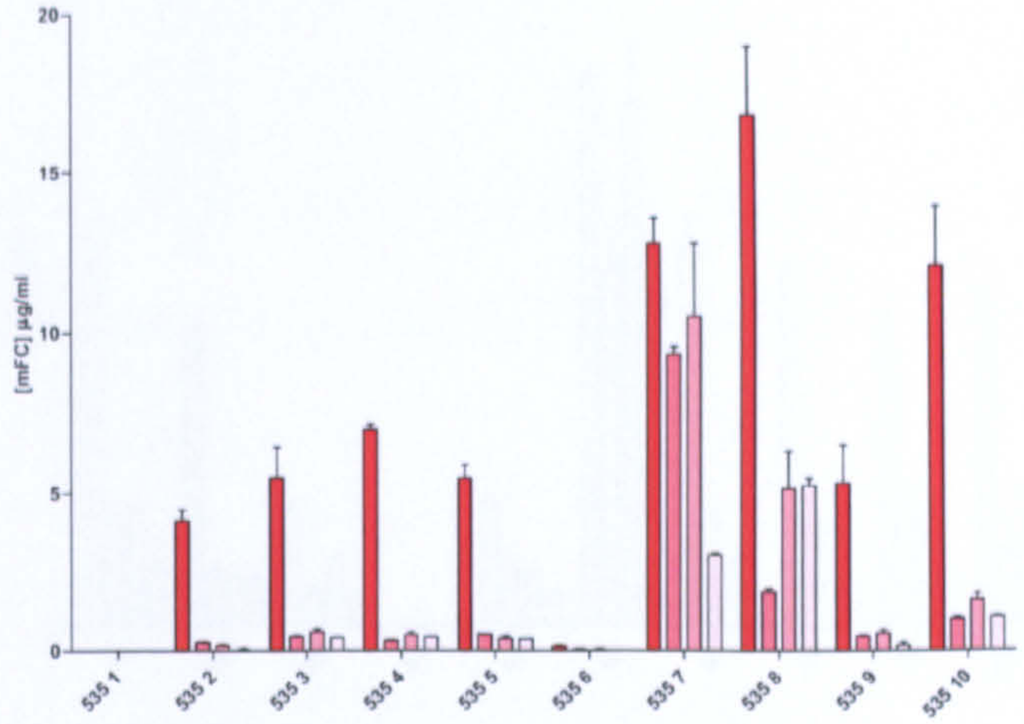
M5, M6, M8, M9 and M10 the expression profile over the 120 generations was different. This was the only MAR X_S29 clone whose expression levels remained stable over the 120 generations with 97% of the starting expression being observed at the end of the stability study.

Expression levels for the STAR 7 clones were the lowest of the elements studied, with antibody titres below the Ab535 control clones. The highest expressing clone was S6 which was 2.5µg/ml. A decline in expression was observed in this clone with only 65% of expression remaining after 120 generations. However, the decrease in expression in the remaining clones, which all shared a similar expression profile over the 120 generations, was more pronounced. The largest decrease in expression was observed after 40 generations. Subsequently, the decline in expression was more gradual. No antibody expression was detected for S4. Although a band was detected in the Southern blot hybridisation (Figure 5.12) no HC or LC DNA was detected by Q-PCR (Figure 5.4) therefore, the lack of antibody expression from this clone was not unexpected.

A decrease in expression levels was observed for all cHS4 tandem clones during the 120 generations. H5 had the smallest decrease with 63% of expression remaining after 120 generations. Somewhat surprisingly, the expression levels for H1 and H2 were higher at 40 generations than at the start of the culture period. This could have been due to differences in cell growth in the 6 well plates. Although cells were seeded at a known cell density, cell counts were not performed through the culture period. Therefore, differences in cell growth in the two cultures could account for the increase in expression. Despite these anomalies there was a decrease in expression

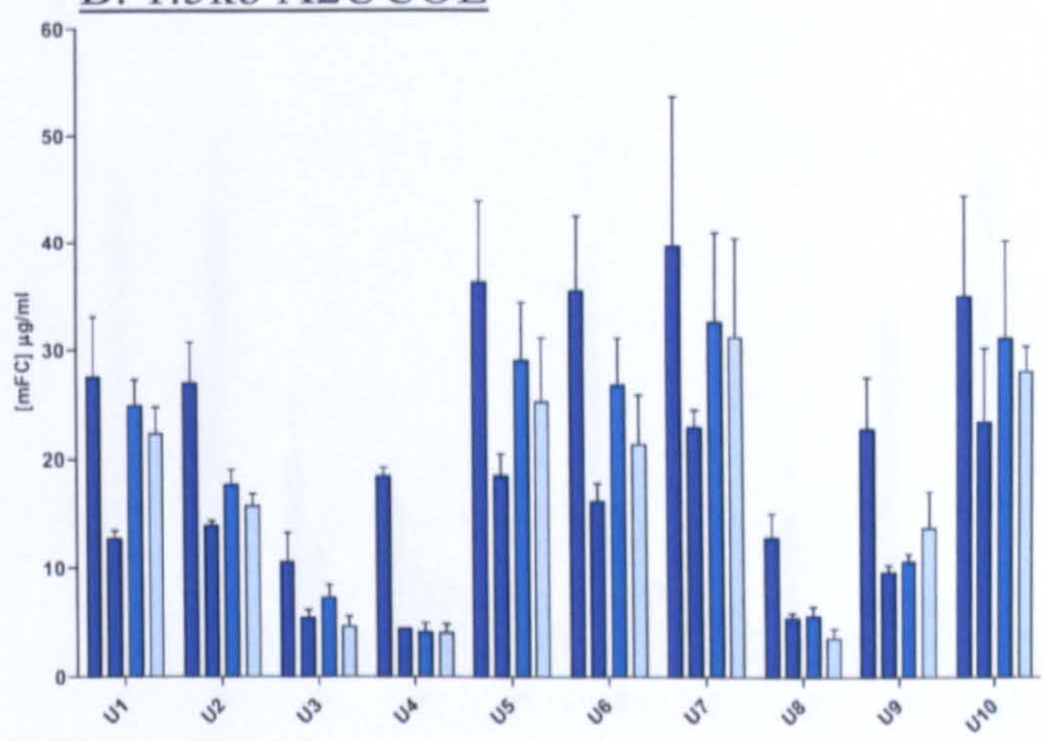
for these two clones over the entire period of the study. For the majority of clones (H3, H4, H6, H7 H8, H9 and H10) the largest decline in expression levels occurred after 40 generations. For the remainder of the culture period further decreases were small.

A. Ab535 clones



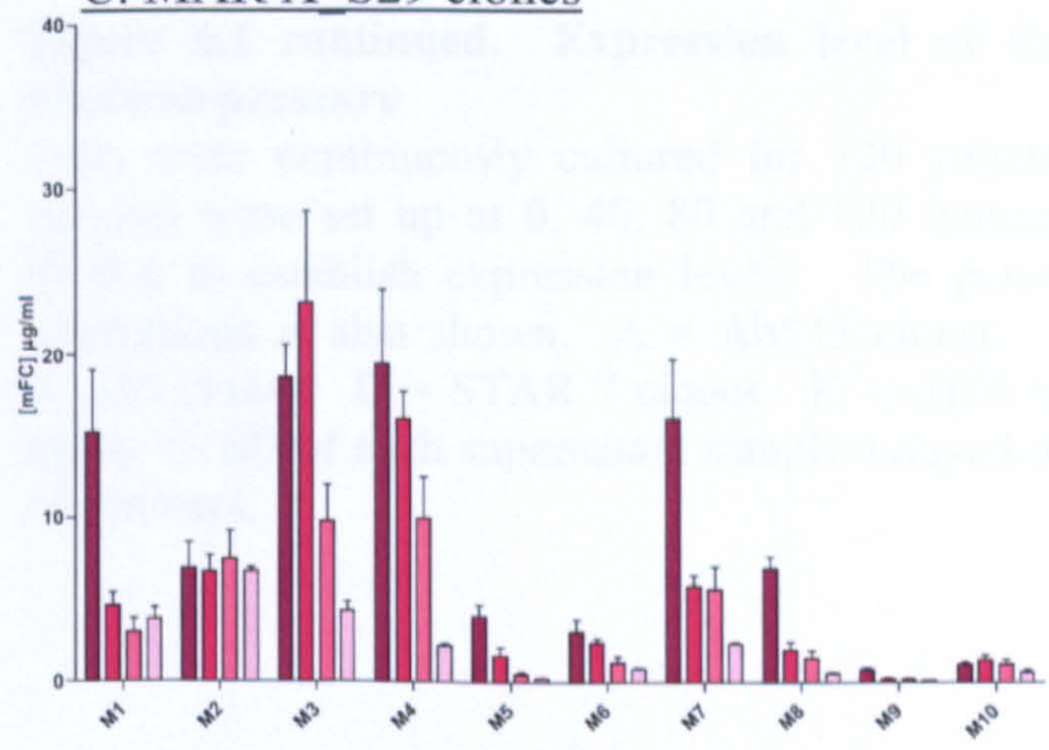
Ab535	% Expression remaining
535 1	0
535 2	1
535 3	7
535 4	6
535 5	7
535 6	0
535 7	24
535 8	31
535 9	3
535 10	9

B. 1.5kb A2UCOE



1.5kb A2UCOE	% Expression remaining
U1	81
U2	58
U3	44
U4	22
U5	69
U6	60
U7	79
U8	28
U9	61
U10	80

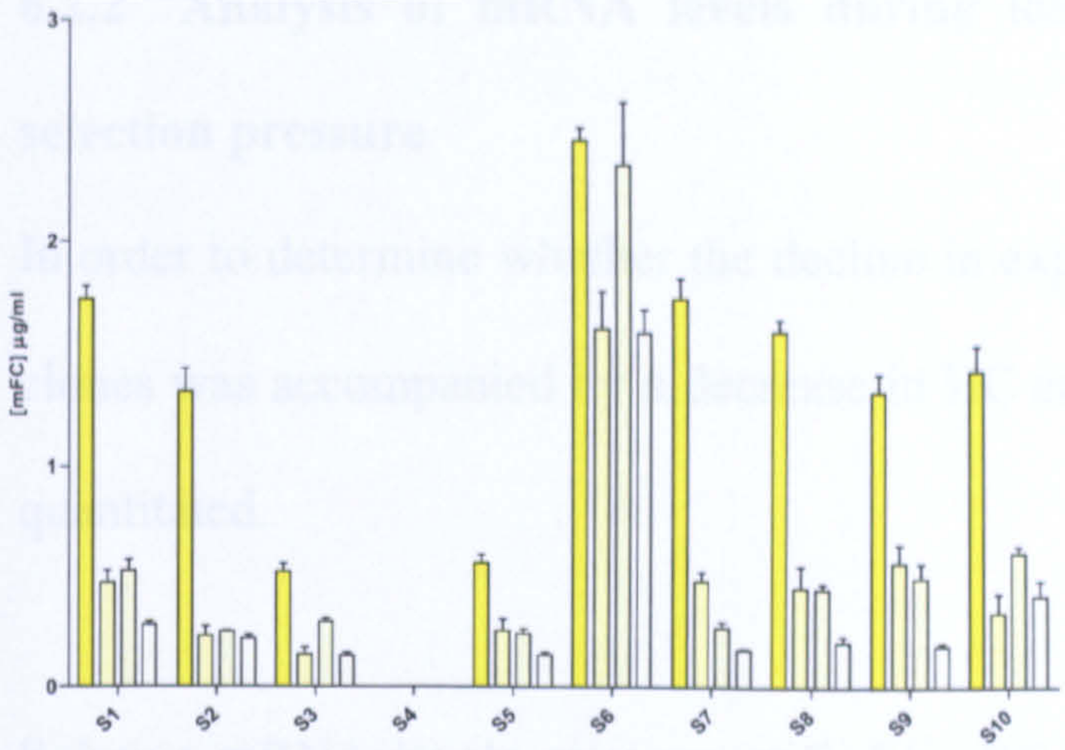
C. MAR X_S29 clones



MAR X_S29	% Expression remaining
M1	25
M2	97
M3	23
M4	11
M5	6
M6	26
M7	15
M8	8
M9	20
M10	58

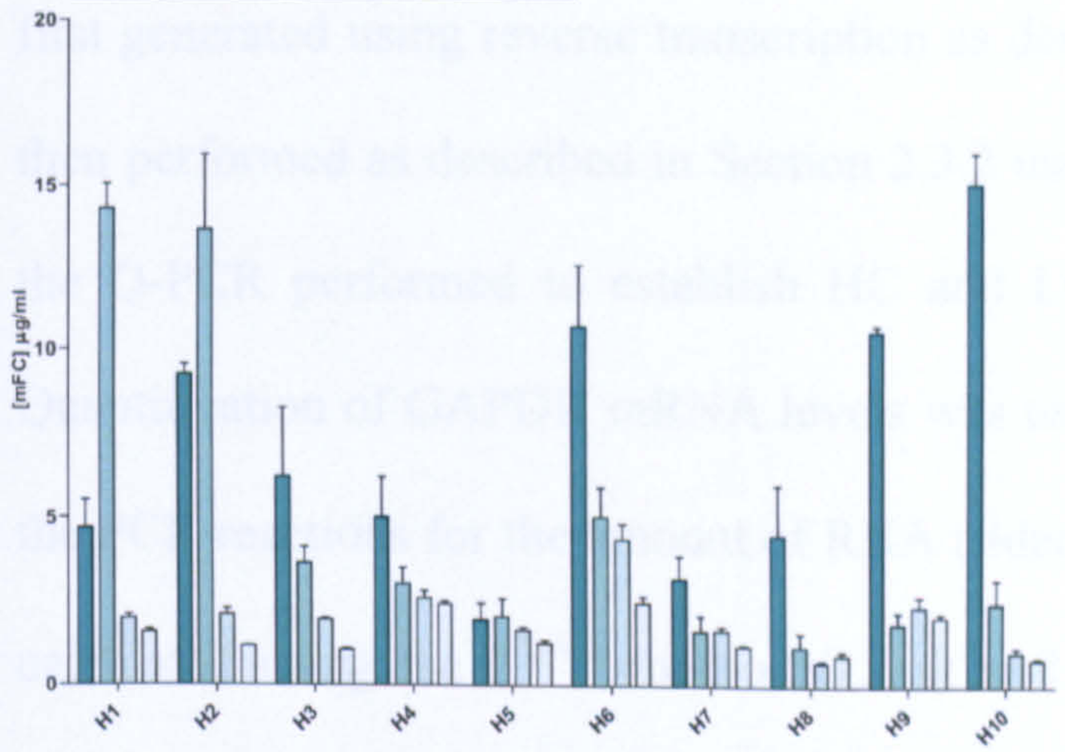
Figure 6.1 Expression level of clones when cultured in the absence of selection pressure
Cells were continuously cultured for 120 generations without MSX. 10 day overgrow cultures were set up at 0, 40, 80 and 120 generations. Supernatant was assayed by mFc ELISA to establish expression levels. The percentage of expression remaining after 120 generations is also shown. A = Ab535 clones. B = 1.5kb A2UCOE clones. C = MAR X_S29 clones. D = STAR 7 clones. E = cHS4 tandem clones. Each data point represents mean +/- SD of each supernatant sample assayed in triplicate wells. Data shown is from one experiment.

D. STAR 7 clones



STAR 7	% Expression remaining
S1	16
S2	17
S3	27
S4	0
S5	26
S6	65
S7	10
S8	13
S9	14
S10	30

E. cHS4 tandem clones



cHS4 tandem	% Expression remaining
H1	33
H2	12
H3	17
H4	48
H5	63
H6	23
H7	37
H8	20
H9	19
H10	5

Figure 6.1 continued. Expression level of clones when cultured in the absence of selection pressure

Cells were continuously cultured for 120 generations without MSX. 10 day overgrow cultures were set up at 0, 40, 80 and 120 generations. Supernatant was assayed by mFc ELISA to establish expression levels. The percentage of expression remaining after 120 generations is also shown. A = Ab535 clones. B = 1.5kb A2UCOE clones. C = MAR X_S29 clones. D = STAR 7 clones. E = cHS4 tandem clones. Each data point represents mean +/- SD of each supernatant sample assayed in triplicate wells. Data shown is from one experiment.

6.2.2 Analysis of mRNA levels during long-term culture in the absence of selection pressure

In order to determine whether the decline in expression observed for a number of the clones was accompanied by a decrease in HC and LC transcripts, mRNA levels were quantitated.

Relative mRNA levels were quantified by RT-Q-PCR. Single stranded cDNA was first generated using reverse transcription as described in Section 2.3.1. Q-PCR was then performed as described in Section 2.3.2 using the same oligonucleotides used in the Q-PCR performed to establish HC and LC copy number (see Section 5.2.1). Quantification of GAPDH mRNA levels was used as an internal control to normalise the PCR reactions for the amount of RNA added. Relative quantities (RQ) were then calculated using the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen, 2001). Firstly, a ΔC_T (cycle threshold) value is calculated by subtracting the C_T value for GAPDH mRNA levels from the C_T value for the gene of interest. One sample is then chosen as a reference against which to calibrate all other samples. The ΔC_T value of the calibrator is then subtracted from the ΔC_T of each sample to generate $\Delta\Delta C_T$. The RQ value is then calculated using $2^{-\Delta\Delta C_T}$.

C_T values for 535 1, 535 7 and 535 8 clones at 0 generations:

Sample	probe	C_T	Avg C_T
535 1	HC	28.24 27.97 28.10	28.10
	LC	23.26 23.26 23.37	23.30
	GAPDH	15.64 15.88 15.88	15.80
535 7 calibrator	HC	16.16 16.15 16.15	16.15
	LC	11.89 11.76 12.32	11.99
	GAPDH	15.54 15.50 15.52	15.52
535 8	HC	18.12 18.17 18.07	18.12
	LC	13.01 13.27 13.30	13.19
	GAPDH	15.25 15.43 15.26	15.31

A. ΔC_T for 535 1 HC samples: 28.10 – 15.80 = **12.31**
ΔC_T for 535 7 HC samples: 16.16 – 15.52 = **0.64**
ΔC_T for 535 8 HC samples: 18.12 – 15.31 = **2.81**

ΔC_T for 535 1 LC samples: 23.30 – 15.80 = **7.50**
ΔC_T for 535 7 LC samples: 11.99 – 15.52 = **- 3.53**
ΔC_T for 535 8 LC samples: 13.19 – 15.31 = **- 2.12**

B. 535 1 $\Delta\Delta C_T$ HC = 12.31 – 0.64 = **11.67**
535 7 $\Delta\Delta C_T$ HC = 0.64 – 0.64 = **0**
535 8 $\Delta\Delta C_T$ HC = 2.81 – 0.64 = **2.17**

535 1 $\Delta\Delta C_T$ LC = 7.50 - -3.53 = **11.03**
535 7 $\Delta\Delta C_T$ LC = -3.53 - -3.53 = **0**
535 8 $\Delta\Delta C_T$ LC = -2.12—3.53 = **1.41**

C. RQ of 535 1 HC to 535 7 HC = $2^{-11.67}$ = **0.0003**
RQ of 535 1 LC to 535 7 LC = $2^{-11.03}$ = **0.0005**

RQ of 535 8 HC to 535 7 HC = $2^{-2.17}$ = **0.22**
RQ of 535 8 LC to 535 7 LC = $2^{-1.41}$ = **0.38**

Calculations:

A. ΔC_T = HC or LC C_T – GAPDH C_T

B. $\Delta\Delta C_T$ = sample ΔC_T – calibrator sample ΔC_T

C. Relative Quantity (RQ) = $2^{-\Delta\Delta C_T}$

Figure 6.2 Example of data used to calculate RQ values for HC and LC mRNA levels Data generated from RT-Q-PCR for three of the Ab535 clones using the calculations A, B and C. Single stranded cDNA generated from RT reaction sampled in triplicate. C_T values calibrated to 535 7.

Samples were taken at 0, 40, 80 and 120 generations to quantify mRNA levels. Data were presented as relative quantities (RQ) compared to the calibrator sample and are

shown in Figure 6.2 The mRNA profiles correlated closely with the antibody expression profiles over the 120 generations (see Figure 6.3 A-E). Where a decrease in antibody expression was observed this was accompanied by a decrease in mRNA levels.

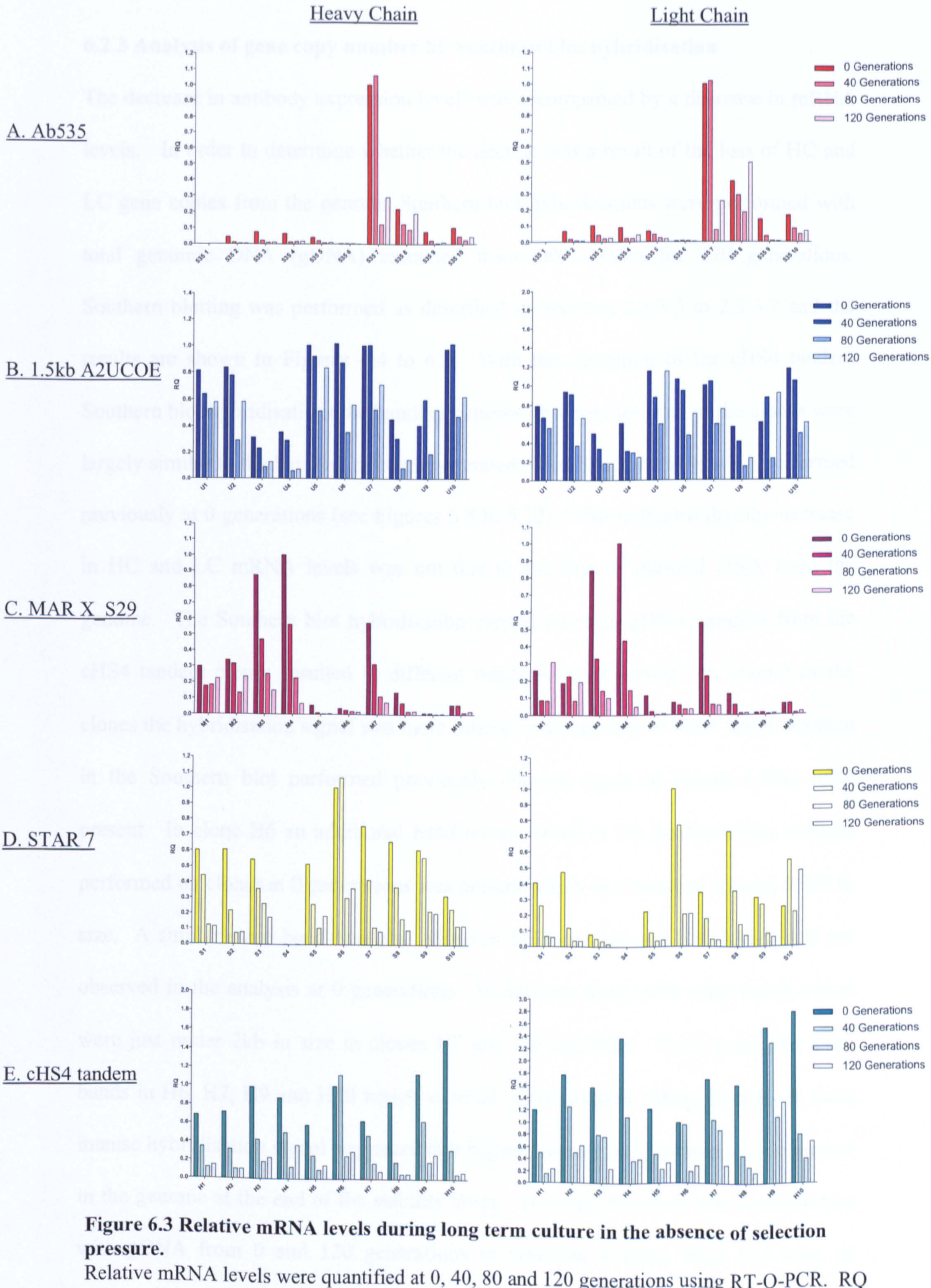


Figure 6.3 Relative mRNA levels during long term culture in the absence of selection pressure.

Relative mRNA levels were quantified at 0, 40, 80 and 120 generations using RT-Q-PCR. RQ values were calculated using the $2^{-\Delta\Delta CT}$ method.

Input RNA quantity was normalised using GAPDH as an internal control. Each data point represents the mean of two independent RT-QPCR assays.

6.2.3 Analysis of gene copy number by Southern blot hybridisation

The decrease in antibody expression levels was accompanied by a decrease in mRNA levels. In order to determine whether the decline was a result of the loss of HC and LC gene copies from the genome Southern blot hybridisations were performed with total genomic DNA (gDNA) extracted from cells grown for 120 generations. Southern blotting was performed as described in Sections 2.3.3.3 to 2.3.3.7 and the results are shown in Figures 6.4 to 6.8. With the exception of the cHS4 tandem Southern blot hybridisations the banding patterns observed for each of the clones were largely similar to the banding patterns generated when Southern blots were performed previously at 0 generations (see Figures 5.8 to 5.12). This indicated that the decrease in HC and LC mRNA levels was not due to the loss of plasmid DNA from the genome. The Southern blot hybridisation carried out with gDNA samples from the cHS4 tandem clones resulted in different bands being observed. In several of the clones the hybridisation signal was more intense and a number of extra bands not seen in the Southern blot performed previously (Shown again in Figure 6.8B.) were present. In clone H6 an additional band not observed in the Southern blot analysis performed on clones at 0 generations was present which was between 3.6 and 4.9kb in size. A similar sized band was also observed in clones H9 and H10 which was not observed in the analysis at 0 generations. In addition there were extra bands which were just under 2kb in size in clones H7 and H9 and H10. There were also extra bands in H6, H7, H9 and H10 which were all approximately 400bp in size. A more intense hybridisation signal suggested that higher plasmid copy numbers were present in the genome at the end of the stability study. Attempts to repeat the Southern blot with gDNA from 0 and 120 generations to establish whether there had been an increase in HC and LC gene copy proved unsuccessful. Therefore, Q-PCR using

gDNA from cells at the end of the stability study was performed in order to determine copy numbers. Figure 6.9 indicates that there had not been a large increase in HC and LC gene copy numbers in the clones at 120 generations. For H6 and H10 the copy numbers were similar at 0 and 120 generations. The Q-PCR results indicated that the copy number of the HC and LC genes for the majority of clones (H1, H2, H3, H4, H5, H7, H8) had increased over the 120 generations, however, it can be seen in Figure 6.9 that the intensity of the highest band in H7 would suggest many more copies in the genome, and the discrepancies in HC and LC gene copy number determined by Q-PCR may be due to the limit of sensitivity of the assay. In contrast to the Southern blot hybridisation, the Q-PCR result for H9 suggested that HC and LC gene copy numbers were lower at 120 generations.

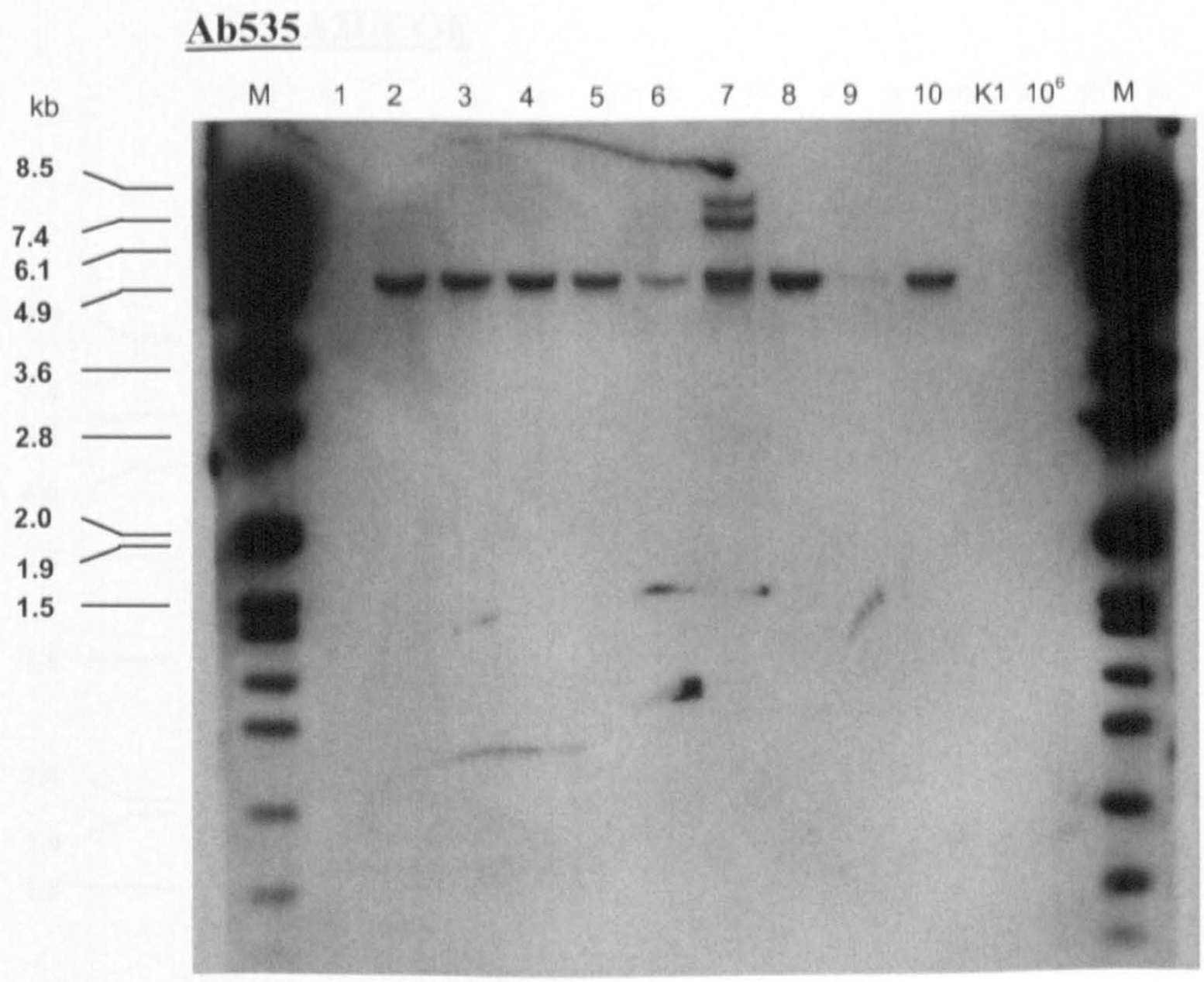


Figure 6.4 Southern blot analysis of Ab535 clones at end of stability study
A. Digested gDNA from cells at 120 generations was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^6 = 1 x 10^6 plasmid copies. 1- 10 = Ab535 clones 535 1 to 535 10.

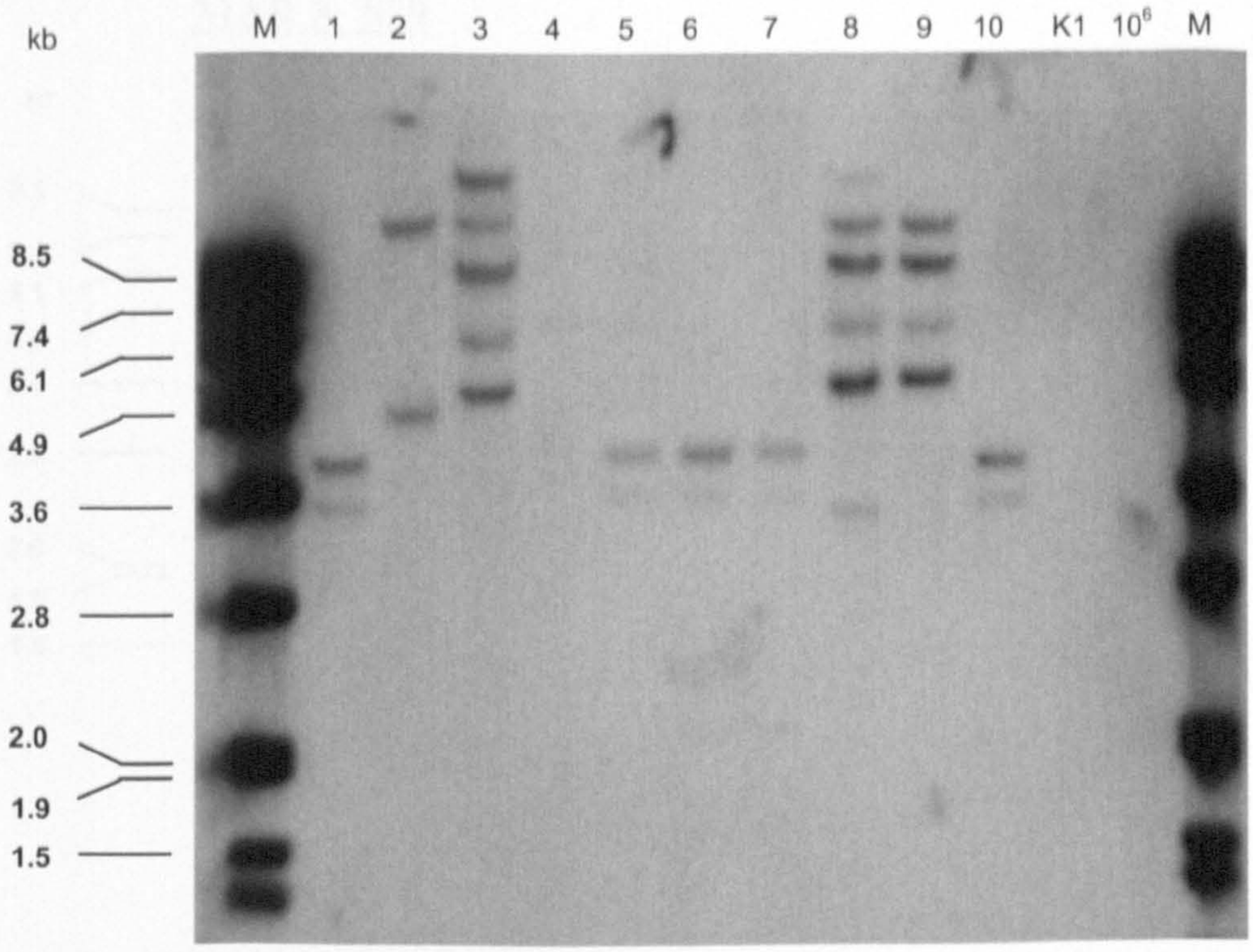


Figure 6.5 Southern blot analysis of 1.5kb A2UCOE clones at end of stability study
A. Digested gDNA from cells at 120 generations was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^6 = 1 x 10^6 plasmid copies. 1- 10 = 1.5kb UCOE clones U1 to U10.

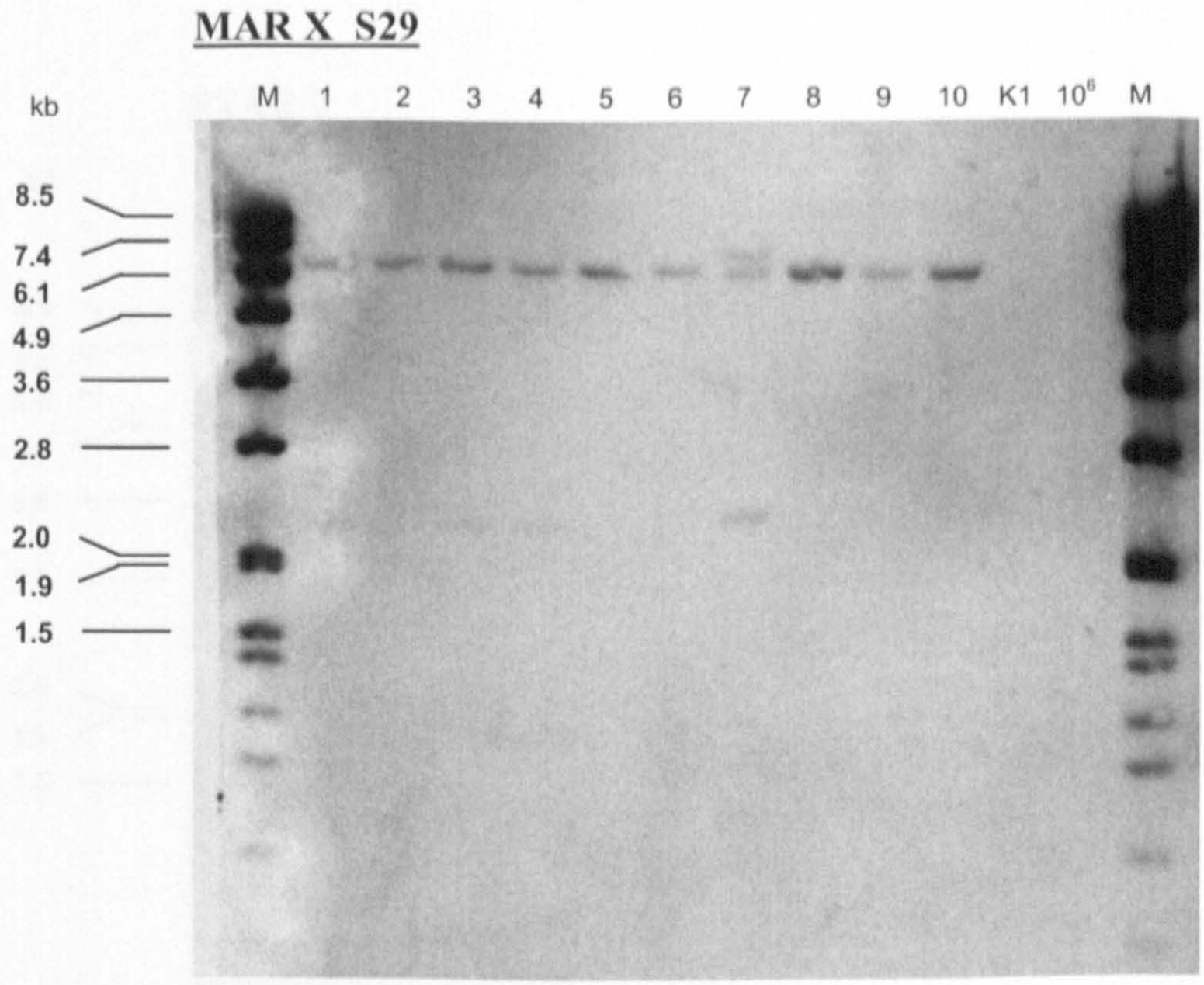


Figure 6.6 Southern blot analysis of MAR X_S29 clones at end of stability study
A. Digested gDNA from cells at 120 generations was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^6 = 1 x 10^6 plasmid copies. 1 - 10 = MAR X_S29 clones M1 to M10.

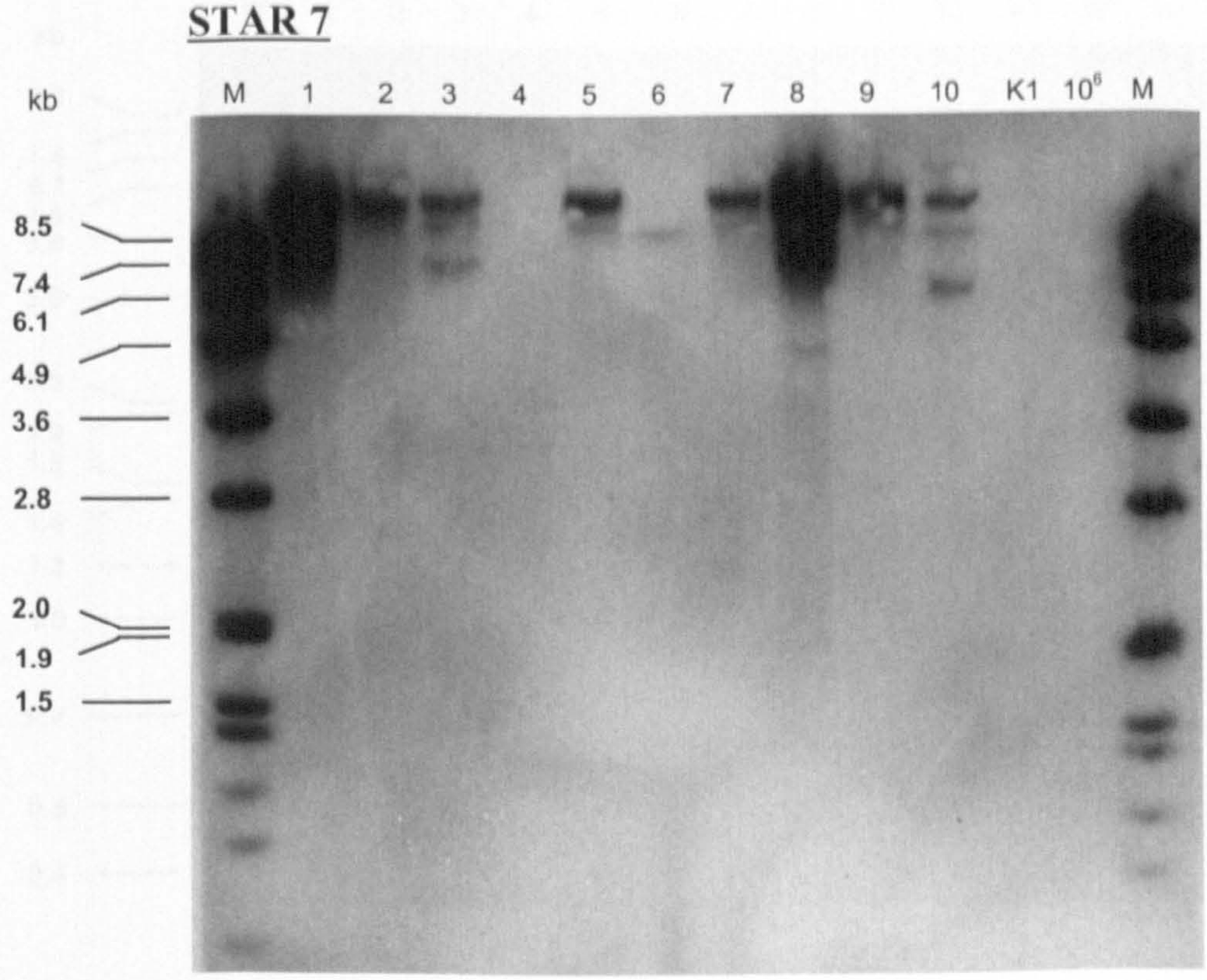
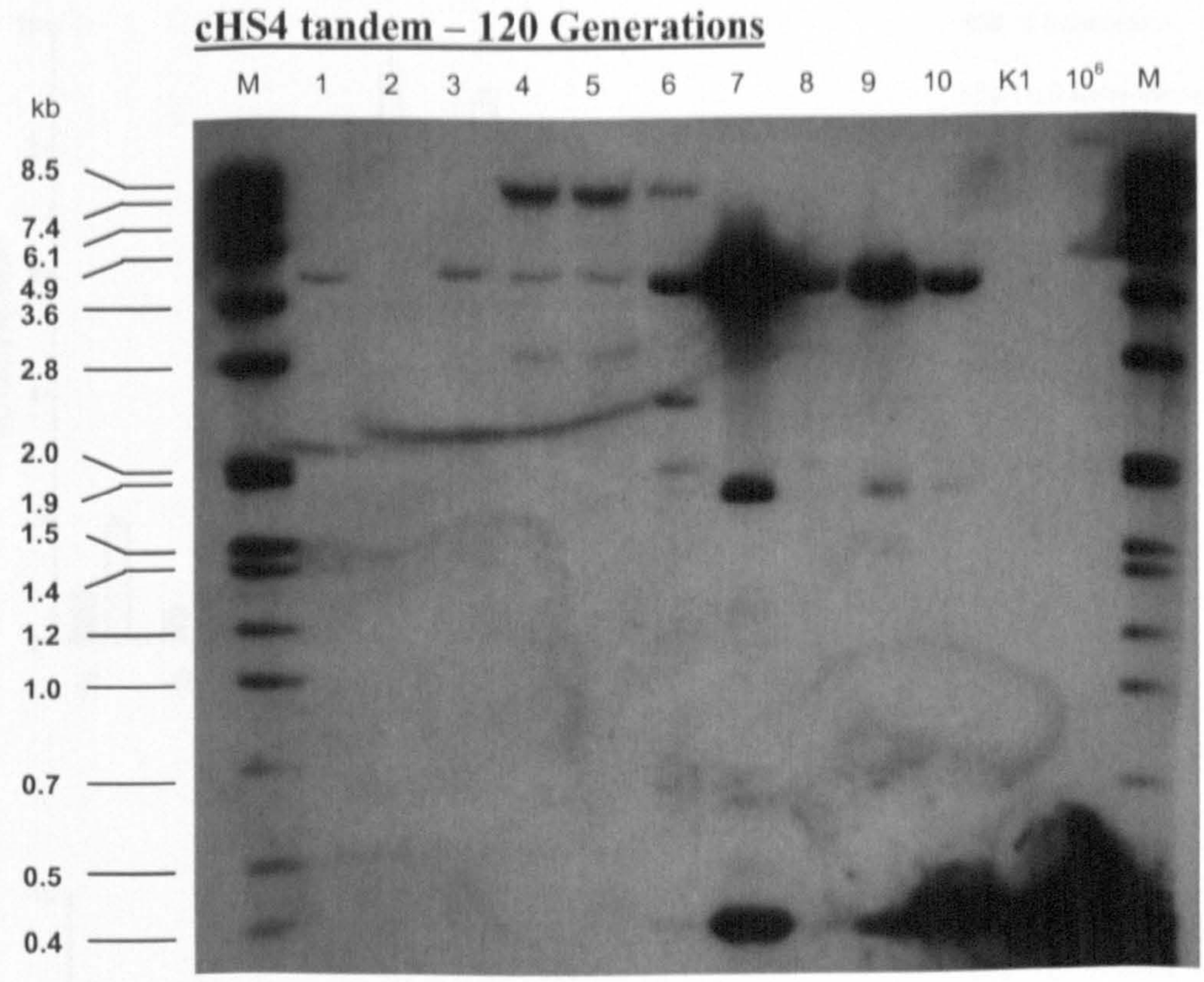


Figure 6.7 Southern blot analysis of STAR 7 clones at end of stability study

A. Digested gDNA from cells at 120 generations was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^6 = 1 x 10^6 plasmid copies. 1 - 10 = STAR 7 clones S1 to S10.

A.



B.

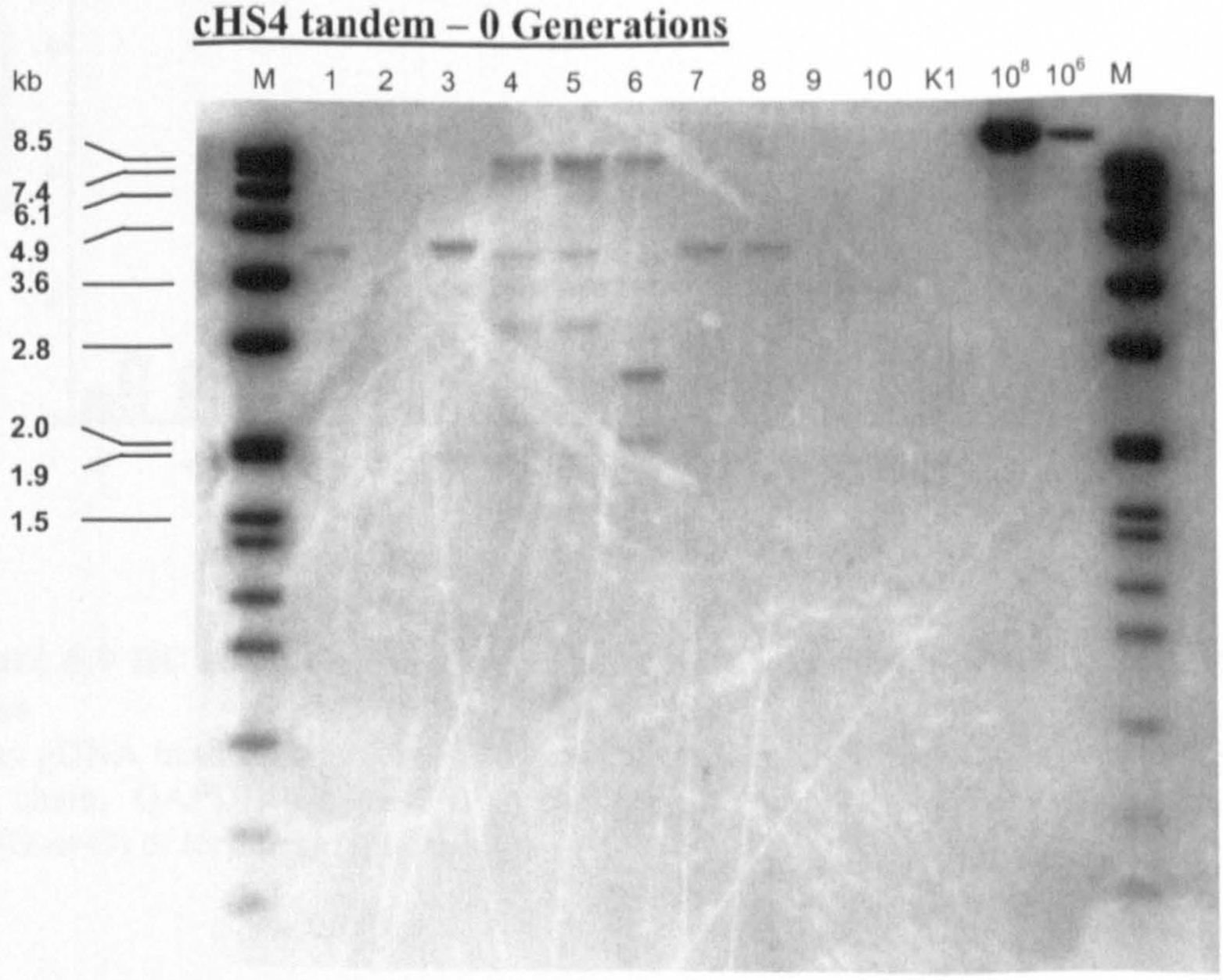


Figure 6.8 Southern blot analysis of cHS4 tandem clones
A. Digested gDNA from cells at 120 generations was analysed by Southern blot hybridisation using a DIG labelled probe. M = DIG labelled DNA ladder. K1 = CHO-K1 gDNA. 10^6 = 1 x 10^6 plasmid copies. 1 - 10 = HS4 tandem clones H1 to H10. B. Southern blot analysis of clones at 0 generations (as shown in Figure 5.12)

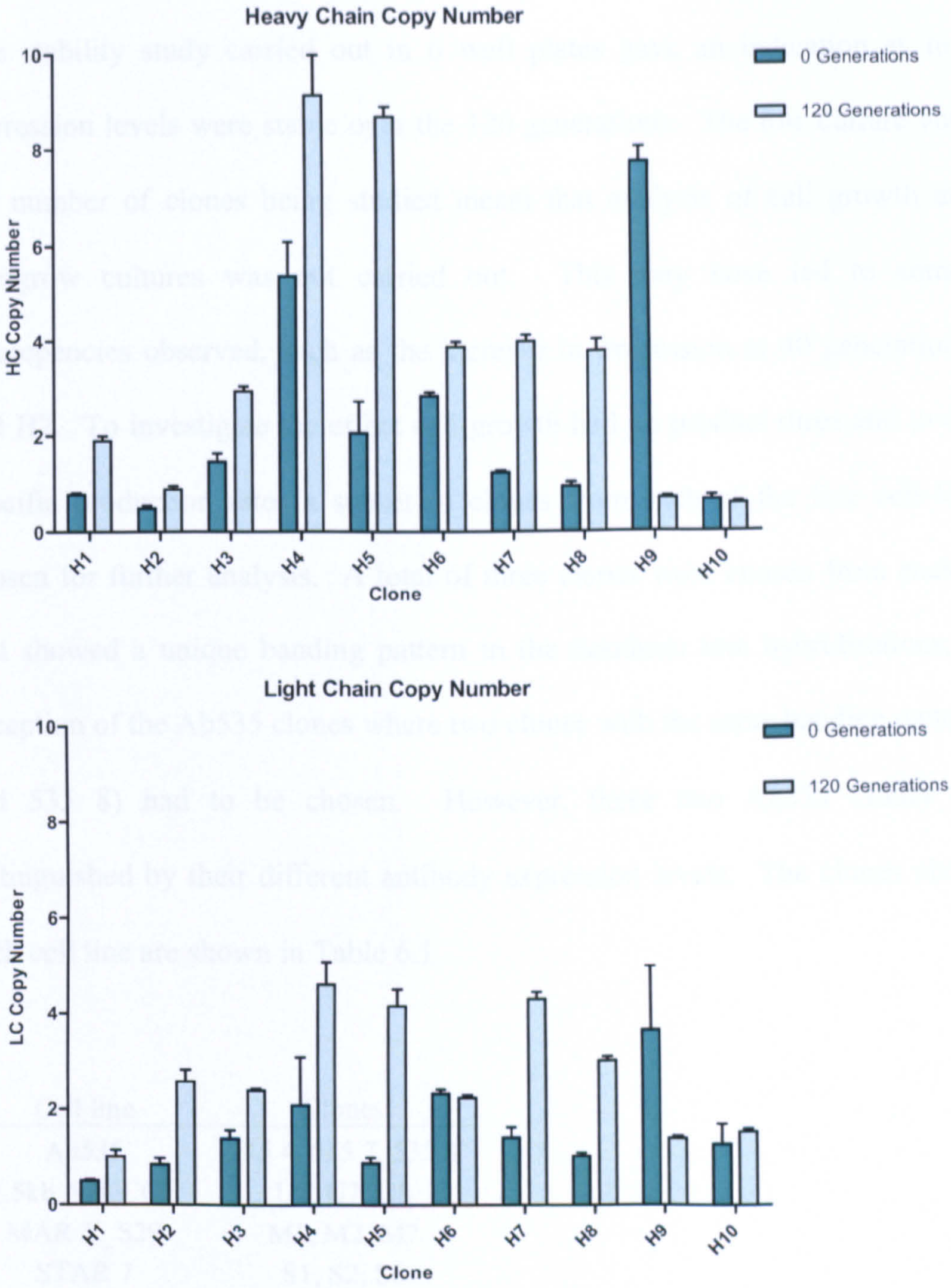


Figure 6.9 HC and LC gene copy number determination by Q-PCR for cHS4 tandem clones

100ng gDNA from cells at 0 and 120 generations was amplified with probes to the heavy and light chain. GAPDH was used as an endogenous control. Each data point represents mean +/- SD (n=3) of technical replicates.

6.2.2 Assessing the specific productivity of cell lines in shake flask culture

The stability study carried out in 6 well plates gave an indication as to whether expression levels were stable over the 120 generations. The low culture volume and the number of clones being studied meant that analysis of cell growth during the overgrow cultures was not carried out. This may have led to some of the discrepencies observed, such as the increase in expression at 40 generations for H1 and H2. To investigate the effect cell growth had on product titres and to determine specific production rates a subset of clones from each of the five cell lines were chosen for further analysis. A total of three clones were chosen from each cell line that showed a unique banding pattern in the Southern blot hybridisations, with the exception of the Ab535 clones where two clones with the same banding pattern (535 4 and 535 8) had to be chosen. However, these two Ab535 clones could be distinguished by their different antibody expression levels. The clones chosen from each cell line are shown in Table 6.1.

Cell line	Clones
Ab535	535 4, 535 7, 535 8
1.5kb A2UCOE	U2, U7, U8
MAR X_S29	M1, M2, M7
STAR 7	S1, S2, S6
cHS4 tandem	H3, H5, H6

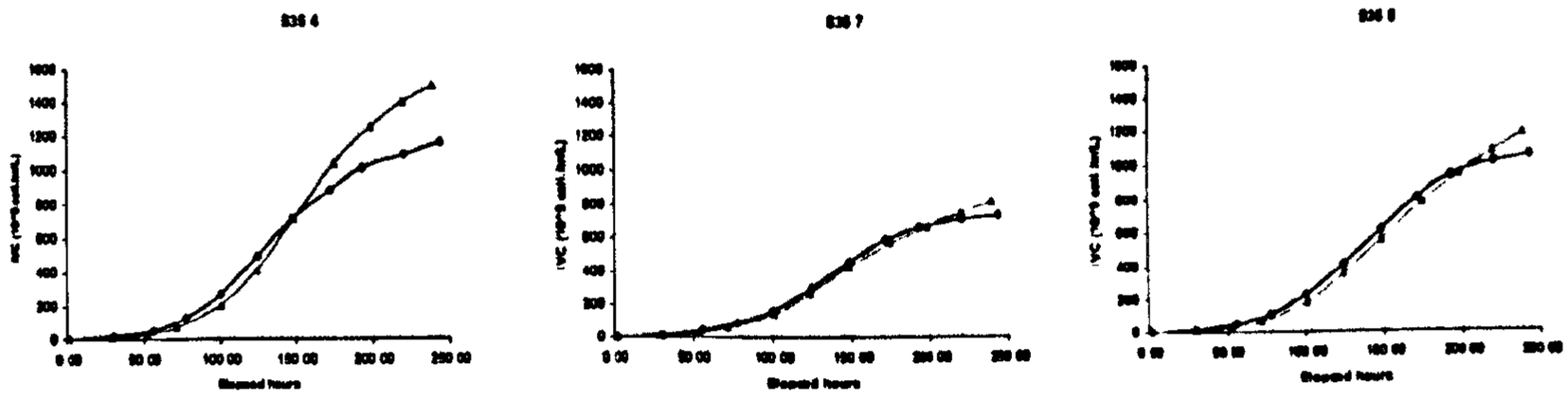
Table 6.1 Clones chosen for specific productivity analysis

Cells from 0 and 120 generations were cultured in 50ml shake flasks. Overgrow cultures were set up with cells seeded at 3×10^5 cells/ml and were grown in batch culture for 10 days. At approximately 24 hour intervals during the culture cell counts were taken and the Integral of Viable Cells (IVC) for each clone were calculated as described in Section 2.4.5.1. Data presented in Figure 6.10 is representative of two

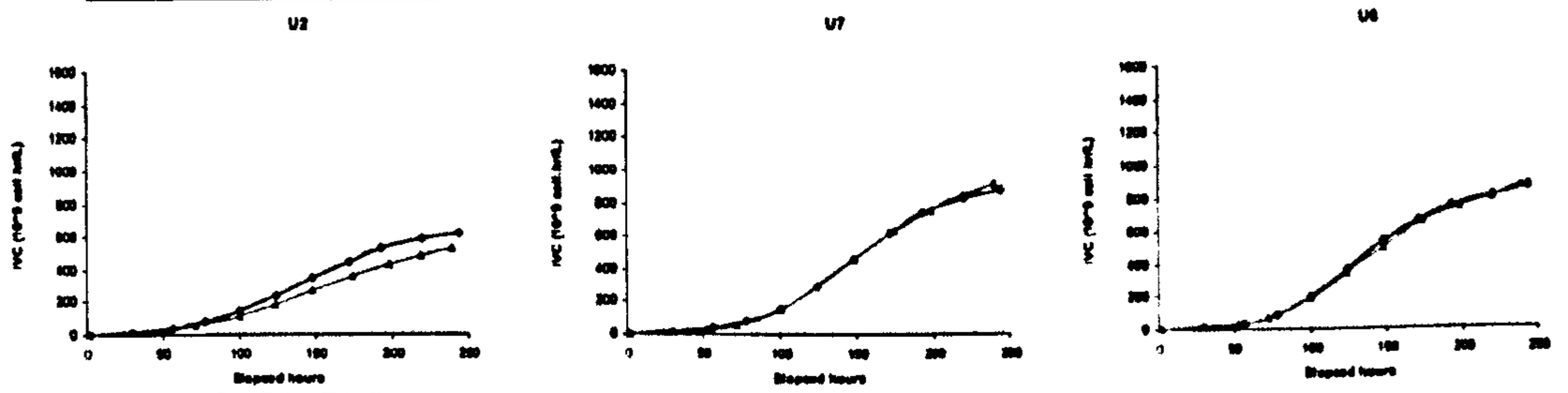
independent overgrow cultures. For the majority of clones, the IVC profiles were similar at 0 and 120 generations. This indicated that cell growth had not altered during the 120 generations. The exception was the two HS4 tandem clones: H5 and H6, where there was almost a two fold increase in the IVC in the overgrow cultures at 120 generations.

There were also differences in the IVCs for the different clones. Clone 535 4 had the highest IVC at 0 generations with over 1100 viable cell hours at the end of the 10 day culture. In contrast the IVC for the three STAR 7 clones were lower with the lowest being S1 which only had an IVC of 434.

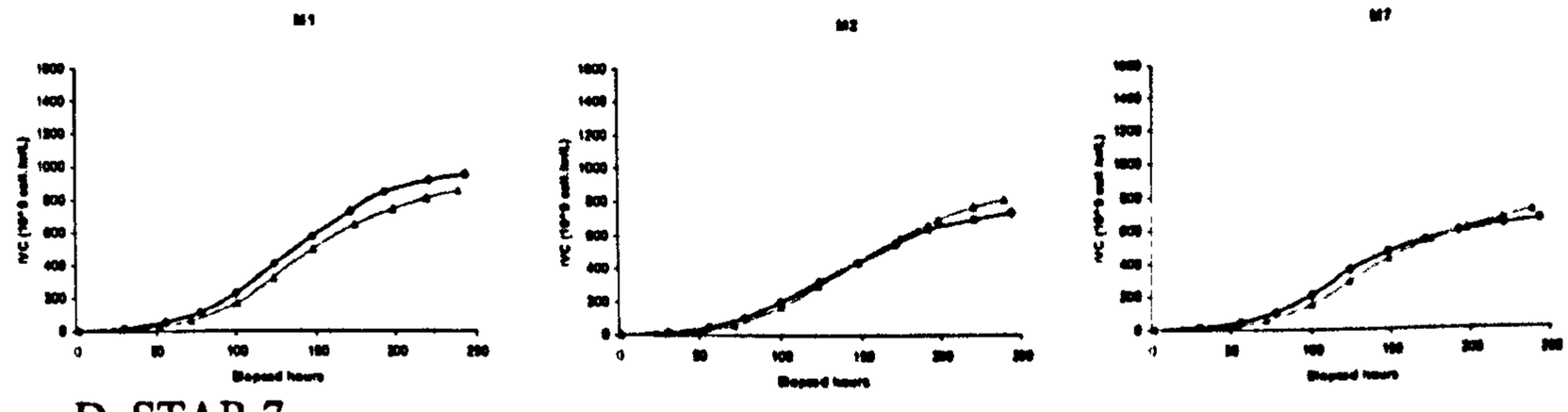
A. Ab535



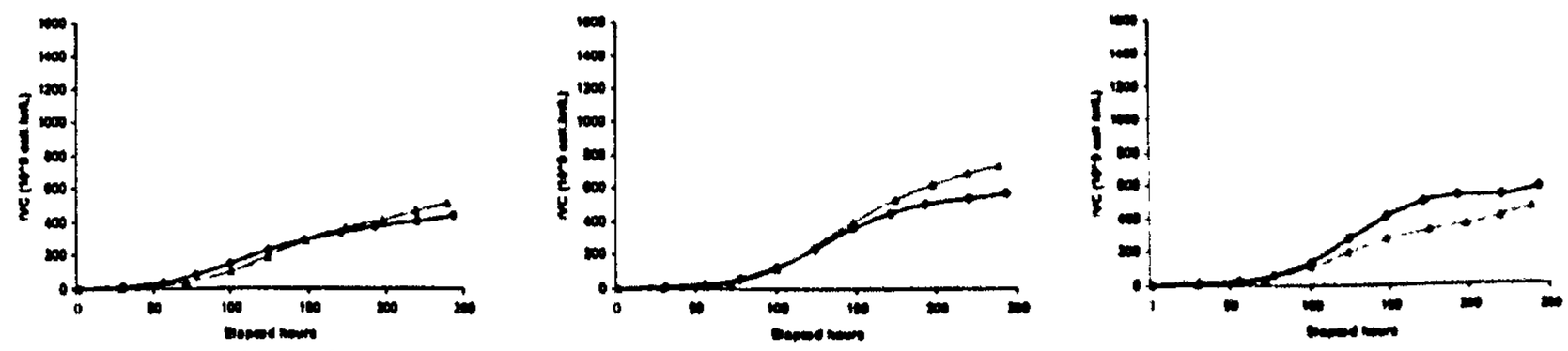
B. 1.5kb A2UCOE



C. MAR X S29



D. STAR 7



E. cHS4 Tandem

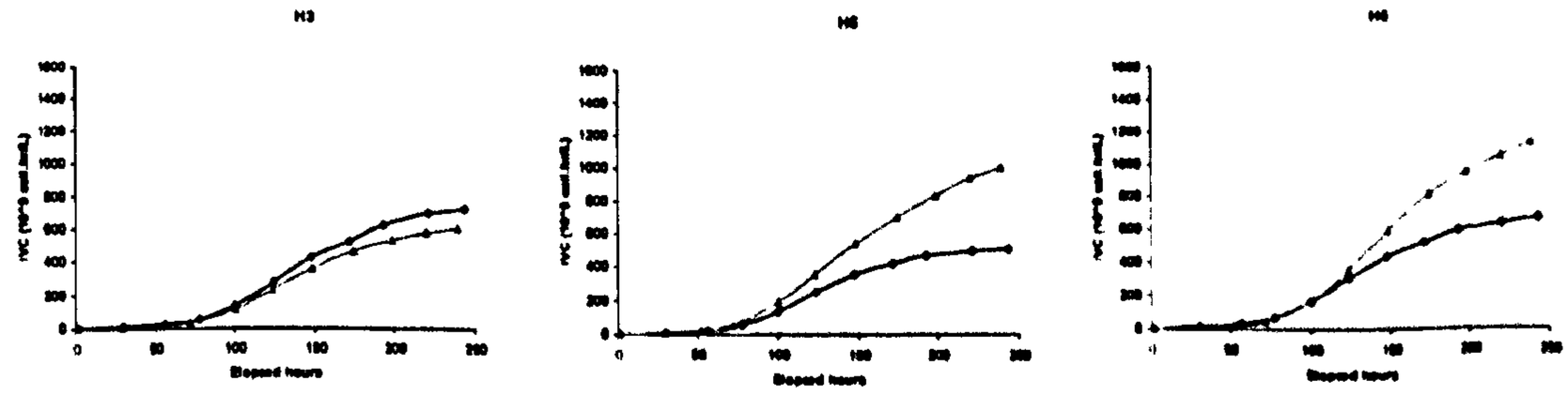
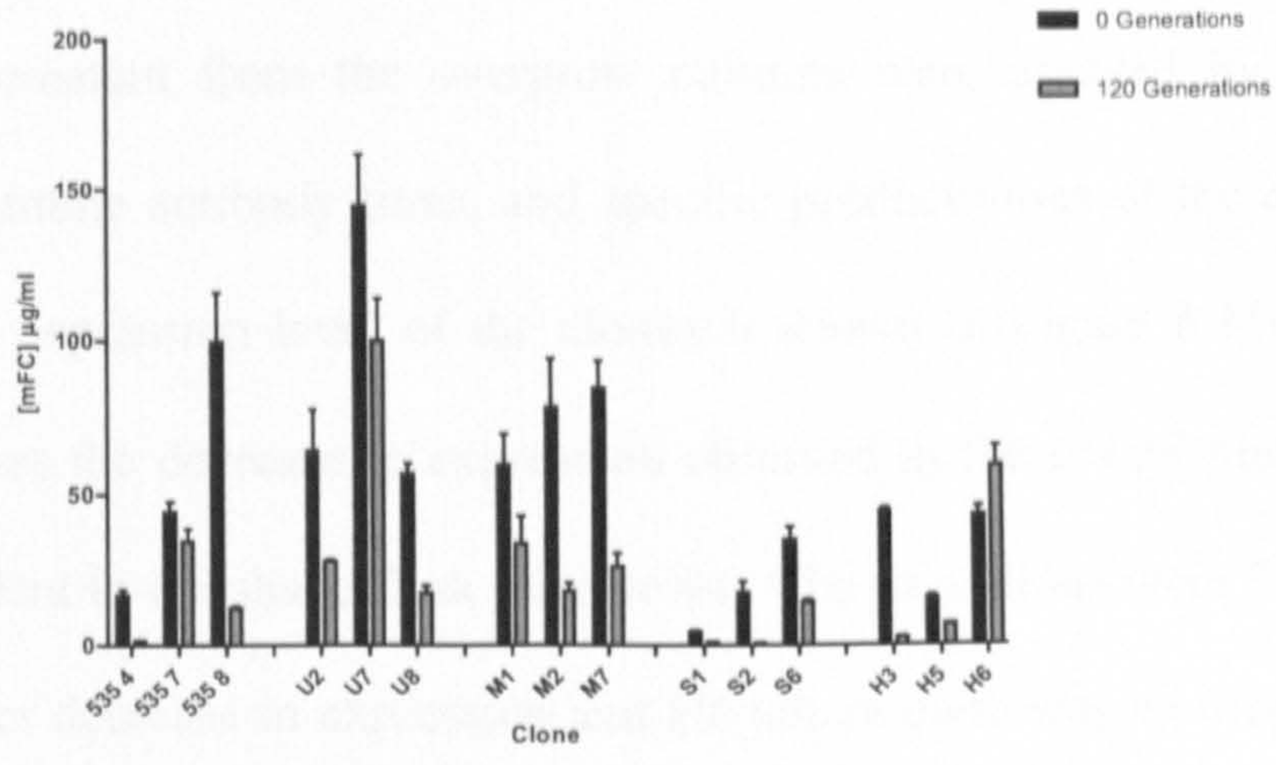
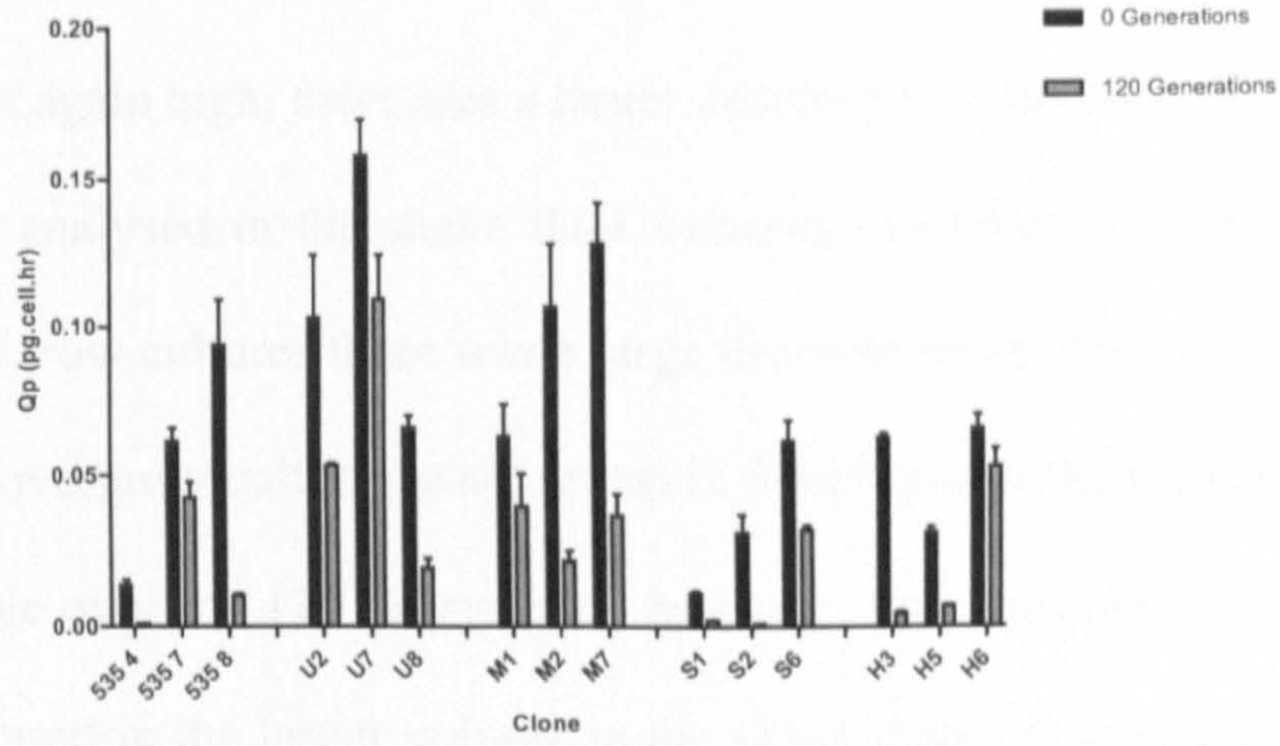


Figure 6.10 Cumulative cell hours of clones at 0 and 120 generations
Cells were seeded at 3 x 10^5 cells/ml in 50 ml shake flask cultures. Cells were grown for 10 days with cell counts taken at 24 hour intervals. The IVC for each clone was calculated as described in Section 2.4.6.2.

A.



B.



Calculating Specific productivity:

$$Q_p = \frac{\text{Antibody titre}}{\text{IVC}}$$

For 535 7: IVC = 723.29

$$Q_p = \frac{44.28}{723.29} = 0.06 \text{ pg.cell.hr}$$

C.

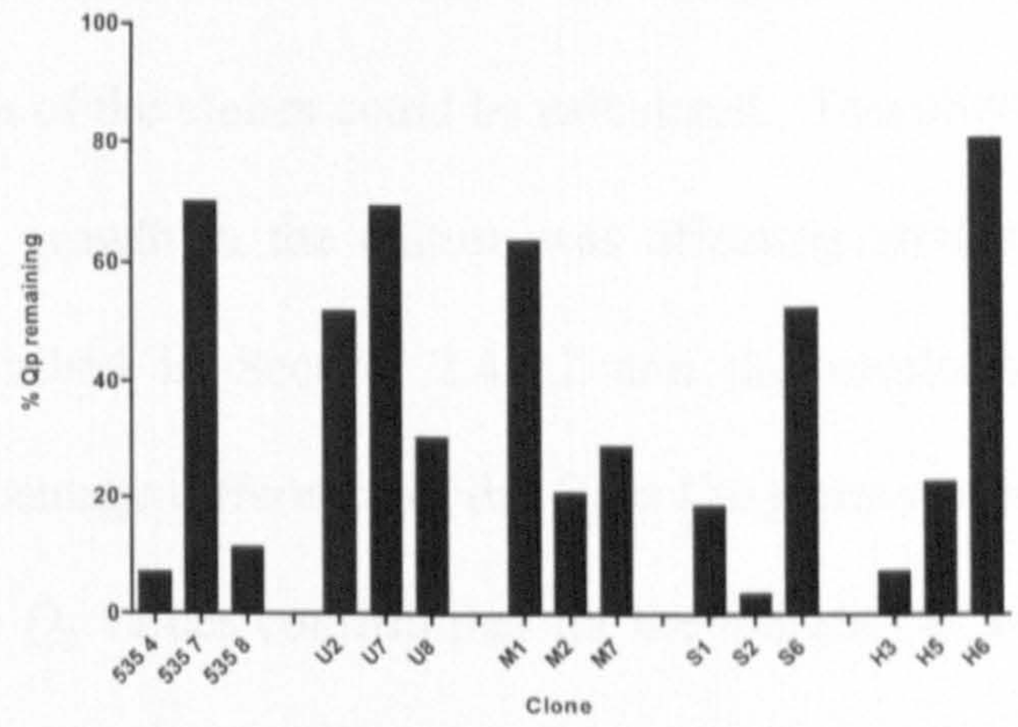


Figure 6.11 Analysis of expression levels and specific productivity of selected clones in shake flask culture at 0 and 120 generations

A subset of clones from each of the five cell lines were analysed in shake flask culture. Cells from 0 and 120 generations were seeded at $3x10^5$ cells/ml in 50ml shake flask culture and grown for 10 days. Viable cell counts were taken at 24 hour intervals during the culture period. Supernatant at day 10 was assayed by a mouse IgG ELISA to determine antibody titre. Each data point represents mean +/- SD (n=3) of technical replicates.
A. Expression levels of clones at 0 and 120 generations. B. Specific productivity (Q_P) of clones at 0 and 120 generations. Example of Q_p calculation for 535 7 shown, Q_p calculated by dividing expression level by IVC. C. Percentage of specific productivity remaining after 120 generations.

Supernatant from the overgrow cultures were assayed by mouse IgG ELISA to determine antibody titres, and specific productivities of the clones were established. The expression level of the clones is shown in Figure 6.11A. For the majority of clones the decrease in expression observed in the 6 well plate overgrows was again evident in the shake flask overgrows. The exceptions were 535 7, which had a much lower decrease in expression and H6 where there was an increase in expression from the culture set up with cells at 120 generations. Although the expression levels for U7 were again high, there was a larger decrease in expression observed when productivity was analysed in the shake flask cultures. In contrast to the results from the 6 well overgrow cultures there was a large decrease in expression levels in clone M2. When the overgrow cultures were set up in 6 well plates the expression from this clone was stable over the 120 generations, however, this was not the case when the cells were cultured in the larger volume in the shake flask. Due to cell counts being performed at 24 hours intervals and IVCs being calculated, the specific productivities (Q_p) for each of the clones could be calculated. This allowed a more in depth analysis of how cell growth in the culture was affecting productivities. The Q_p was calculated as described in Section 2.4.5.2 and the results are shown in Figure 6.11B. The percentage difference of the Q_p at 120 generations is shown in Figure 6.11C.

The Q_p values confirm that for the majority of clones there was a loss of expression during the long term culture in the absence of selection. The difference in expression levels between 0 and 120 generations for clone H6 would suggest that there was an increase in expression during the long term culture. However, by taking into account the increase in IVC observed for this clone (see Figure 6.10) the specific production rate for the clone at 0 and 120 generations was similar.

6.2.5 Investigating the effect of DNA methylation levels on the regulation of gene expression

DNA methylation is an important epigenetic mark and is involved in the regulation of gene expression. High levels of DNA methylation are associated with gene repression and a condensed chromatin state (Bird and Wolffe, 1999). The data presented in Sections 6.2.1 to 6.2.3 indicates that there was a decrease in expression levels, which was not due to the loss of the integrated plasmid from the genome. This would suggest that the decline in antibody expression was a result of epigenetic regulation, in which DNA methylation may be involved.

In order to determine DNA methylation patterns bisulphite conversion followed by sequencing of the hCMV-MIE promoter regions was performed. When DNA is treated with bisulphite cytosine residues are converted to uracils. However 5-methylcytosine residues remain unchanged (Figure 6.12). The region of interest is amplified by PCR and the products are cloned and sequenced. Cytosine residues in the sequence represent methylated cytosines, while those that are read as thymines represent unmethylated cytosines in the gDNA. The cloning of amplified PCR products allows the methylation status of single DNA strands from individual DNA molecules in the original gDNA sample to be established.

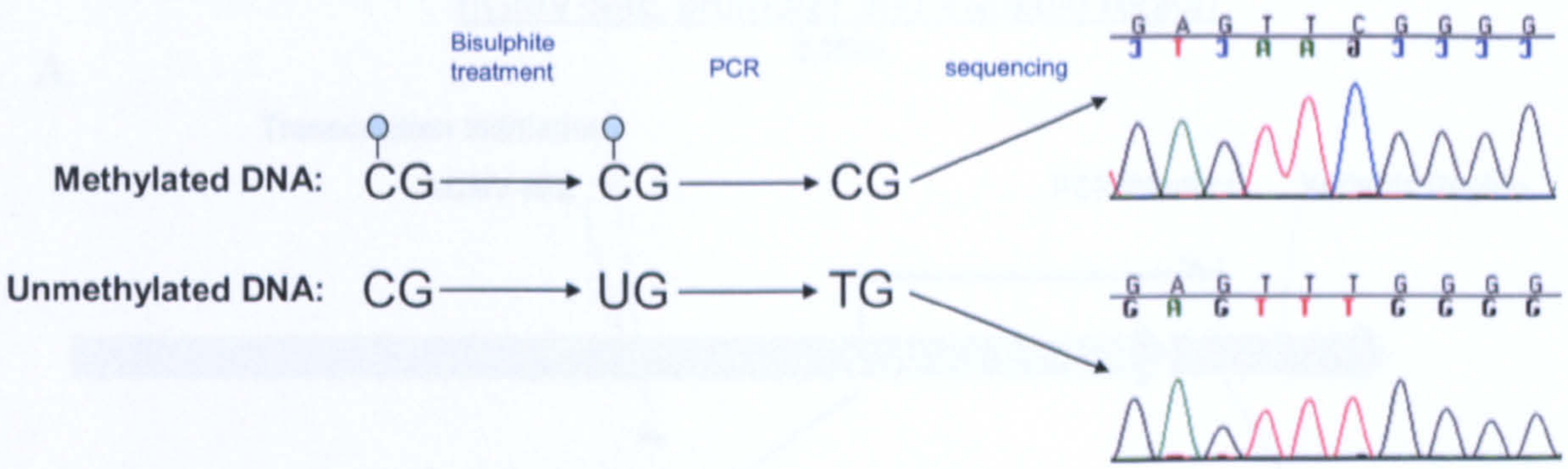


Figure 6.12 Bisulphite sequencing
After treatment with bisulphite methylated cytosine residues remain as cytosine. Unmethylated cytosines are converted to uracils and then thymidine after PCR amplification. Sequencing of PCR products allows methylated and unmethylated cytosine residues to be distinguished.

gDNA samples from the start and end of the stability study (0 and 120 generations) were analysed to investigate whether DNA methylation levels increased during the long-term culture period. The three clones from each of the five cell lines indicated in Table 6.1 were chosen for the bisulphite sequencing.

In order to distinguish between the HC and LC hCMV-MIE promoters DNA oligonucleotides were designed to amplify the 3' region of the hCMV-MIE promoter and the 5' region of the VL or VH region as indicated in Figure 6.13. The amplified PCR products were 660bp for the LC promoter region and 700bp for the HC promoter region which allowed a total of 37 and 39 CpG dinucleotides to be analysed in the LC and HC promoter regions respectively.

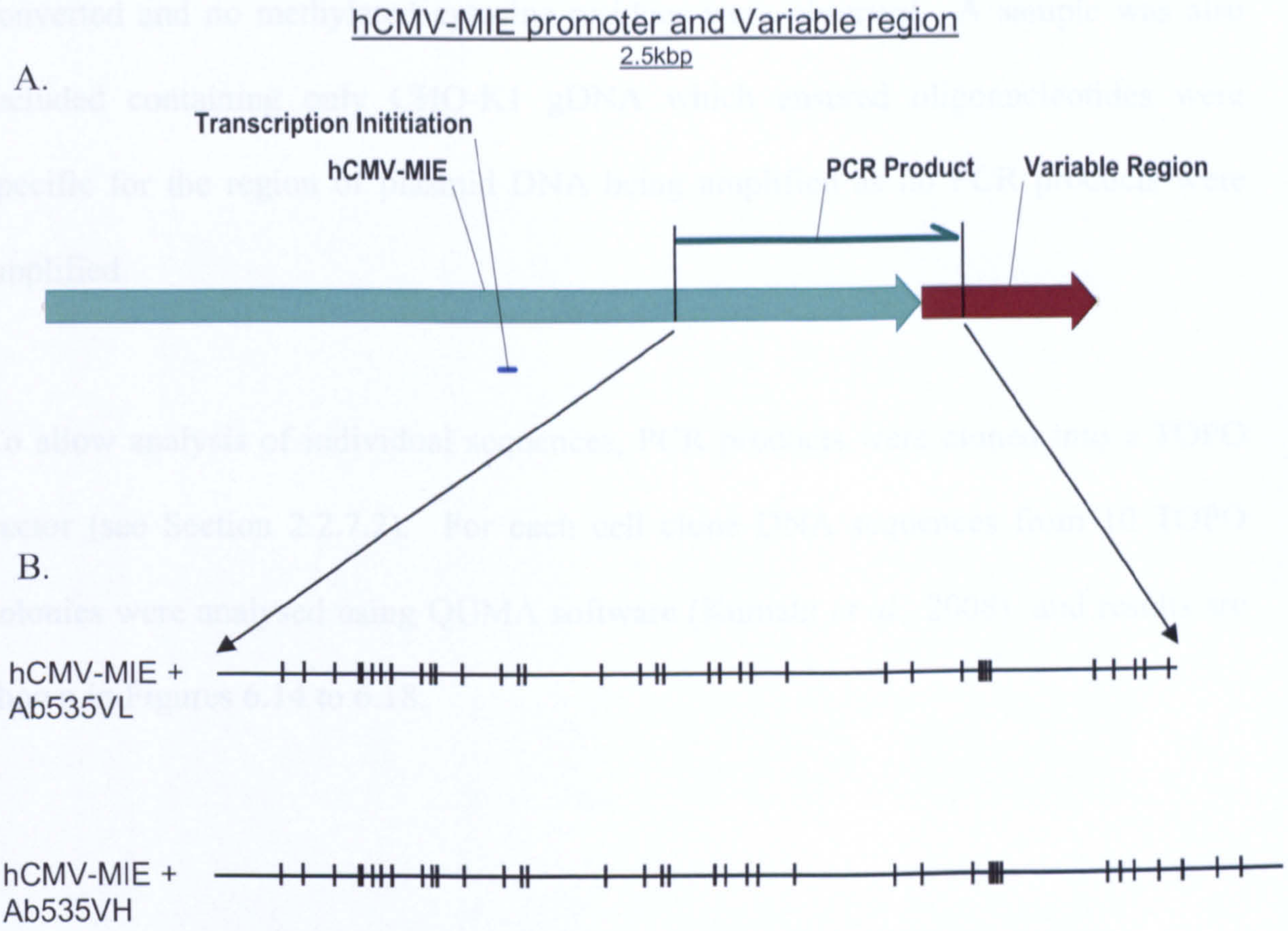


Figure 6.13 Illustration of the plasmid region amplified after bisulphite conversion and position of CpG dinucleotides

A. Location of region within hCMV-MIE promoter and variable region amplified by PCR. LC PCR product = 660bp, HC PCR product = 700bp. B. Position of CpG sites within PCR product shown by vertical lines CpG sites in LC region = 37, CpG sites in HC region = 39.

Oligonucleotides were designed using the Methyl Primer Express software (Applied Biosystems) which ensured no CpG dinucleotides were present in the oligonucleotides themselves (oligonucleotides 5.1 to 5.8 in Appendix 2). If CpG dinucleotides were present in the oligonucleotide differential amplification of the region would occur depending on levels of methylation. Nested PCR was performed to obtain sufficient amounts of amplified product.

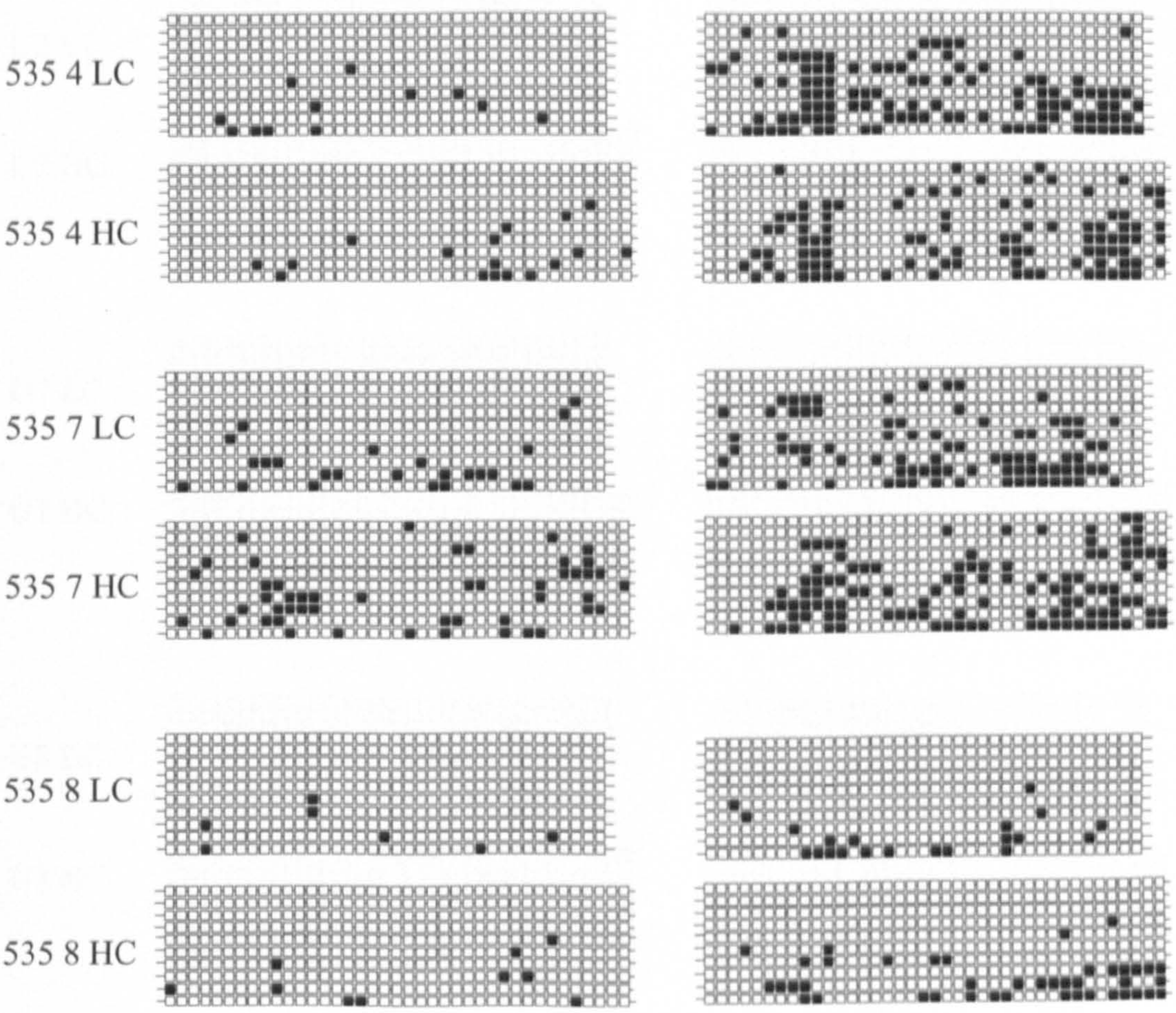
To ensure complete conversion of the cytosines was taking place a control reaction was included which consisted of CHO-K1 gDNA spiked with 1 x 10^6 copies of plasmid DNA spiked in. Following bisulphite conversion all cytosine residues were

converted and no methylated cytosine residues were observed. A sample was also included containing only CHO-K1 gDNA which ensured oligonucleotides were specific for the region of plasmid DNA being amplified as no PCR products were amplified.

To allow analysis of individual sequences, PCR products were cloned into a TOPO vector (see Section 2.2.7.2). For each cell clone DNA sequences from 10 TOPO colonies were analysed using QUMA software (Kumaki *et al.*, 2008) and results are shown in Figures 6.14 to 6.18.

Ab535 clones

A.



B.

	Light Chain		Heavy Chain	
	0	**120**	**0**	**120**
535 4	3.2	30.5	4.4	24.6
535 7	6.5	21.8	12.8	28.2
535 8	1.9	5	2.6	10.8

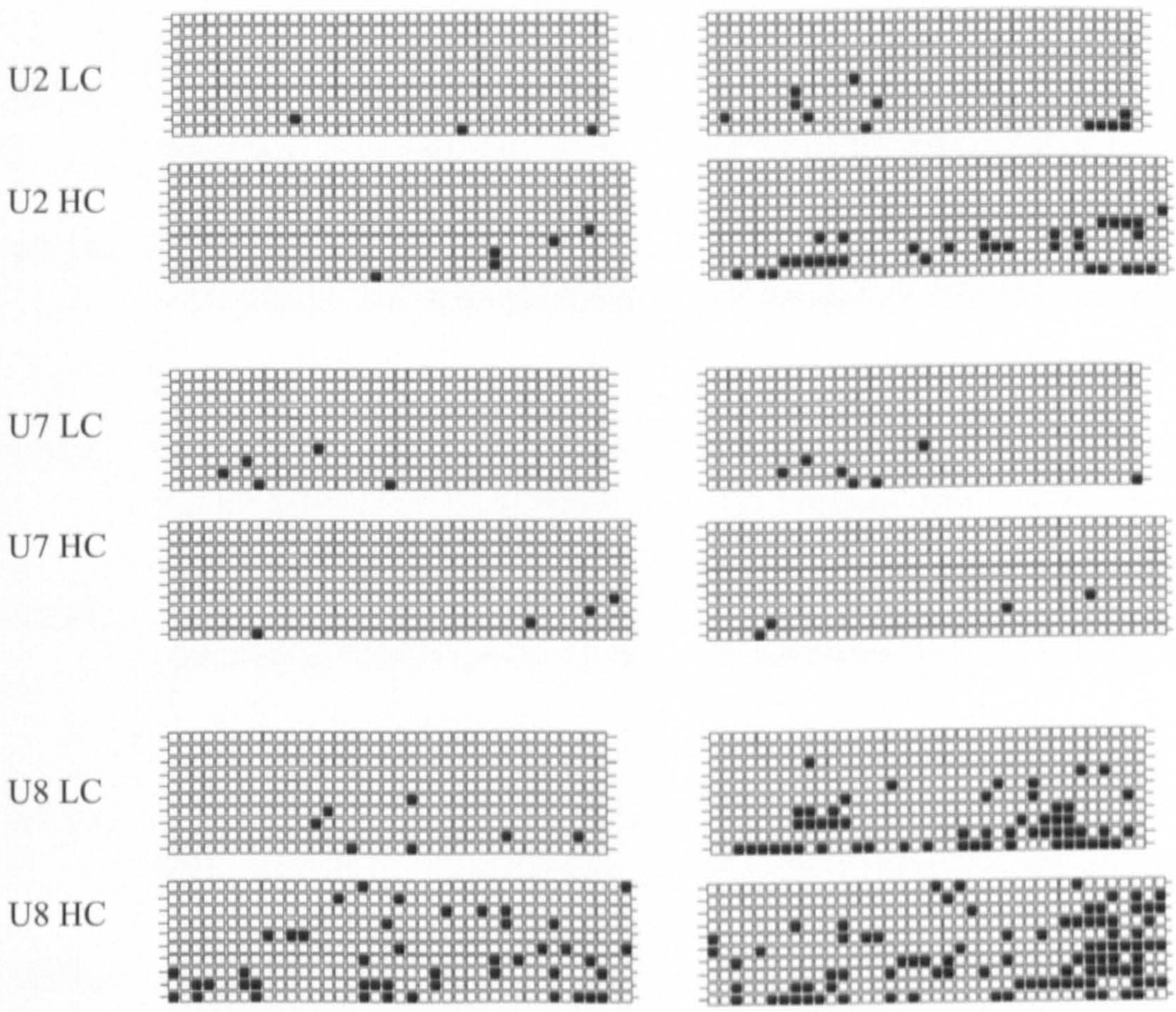
Figure 6.14 Methylation status of CpG dinucleotides within the HC and LC promoter regions in Ab535 clones

Bisulphite sequencing was performed to establish the number of methylated cytosine residues in CpG dinucleotides within the hCMV-MIE promoter. A. Each line represents the result within individual clones of amplified PCR product. White squares = unmethylated CpG dinucleotide, black squares = methylated CpG dinucleotide.

B. Percentage of methylated CpG sites in light and heavy chain promoter region at 0 and 120 generations.

1.5kb A2UCOE clones

A.



B.

	Light Chain		Heavy Chain	
	0	**120**	**0**	**120**
U2	0.8	3.3	1.3	8.5
U7	1.4	1.9	1	1
U8	1.9	14.9	11.5	23.6

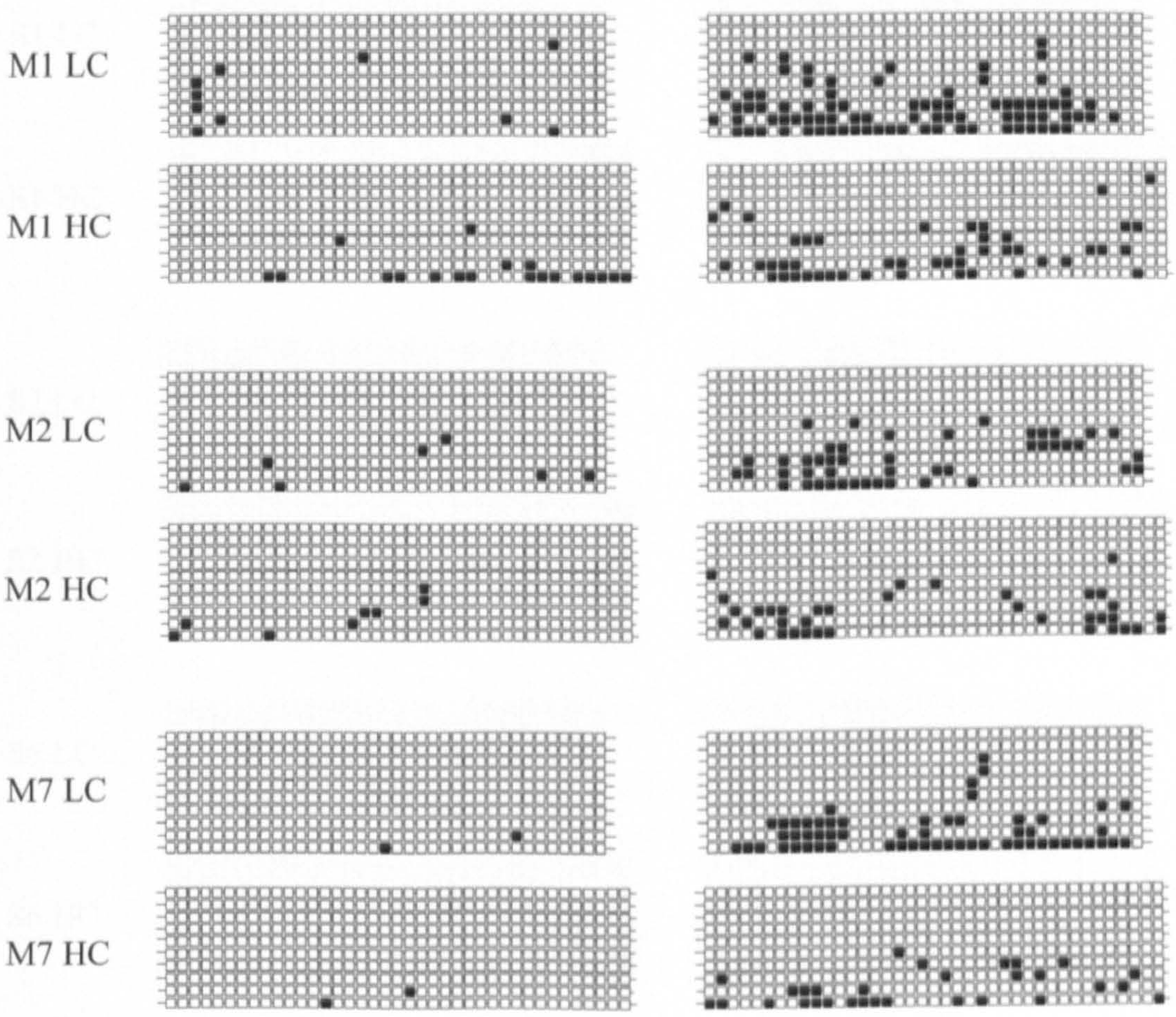
Figure 6.15 Methylation status of CpG dinucleotides within the HC and LC promoter regions in 1.5kb A2UCOE clones

Bisulphite sequencing was performed to establish the number of methylated cytosine residues in CpG dinucleotides within the hCMV-MIE promoter. A. Each line represents the result within individual clones of amplified PCR product. White squares = unmethylated CpG dinucleotide, black squares = methylated CpG dinucleotide.

B. Percentage of methylated CpG sites in light and heavy chain promoter region at 0 and 120 generations.

MAR X_S29 clones

A.



B.

	Light Chain		Heavy Chain	
	0	**120**	**0**	**120**
M1	2.7	22.4	5.9	12.1
M2	1.9	12.5	2.1	10.5
M7	0.5	16	0.6	8.2

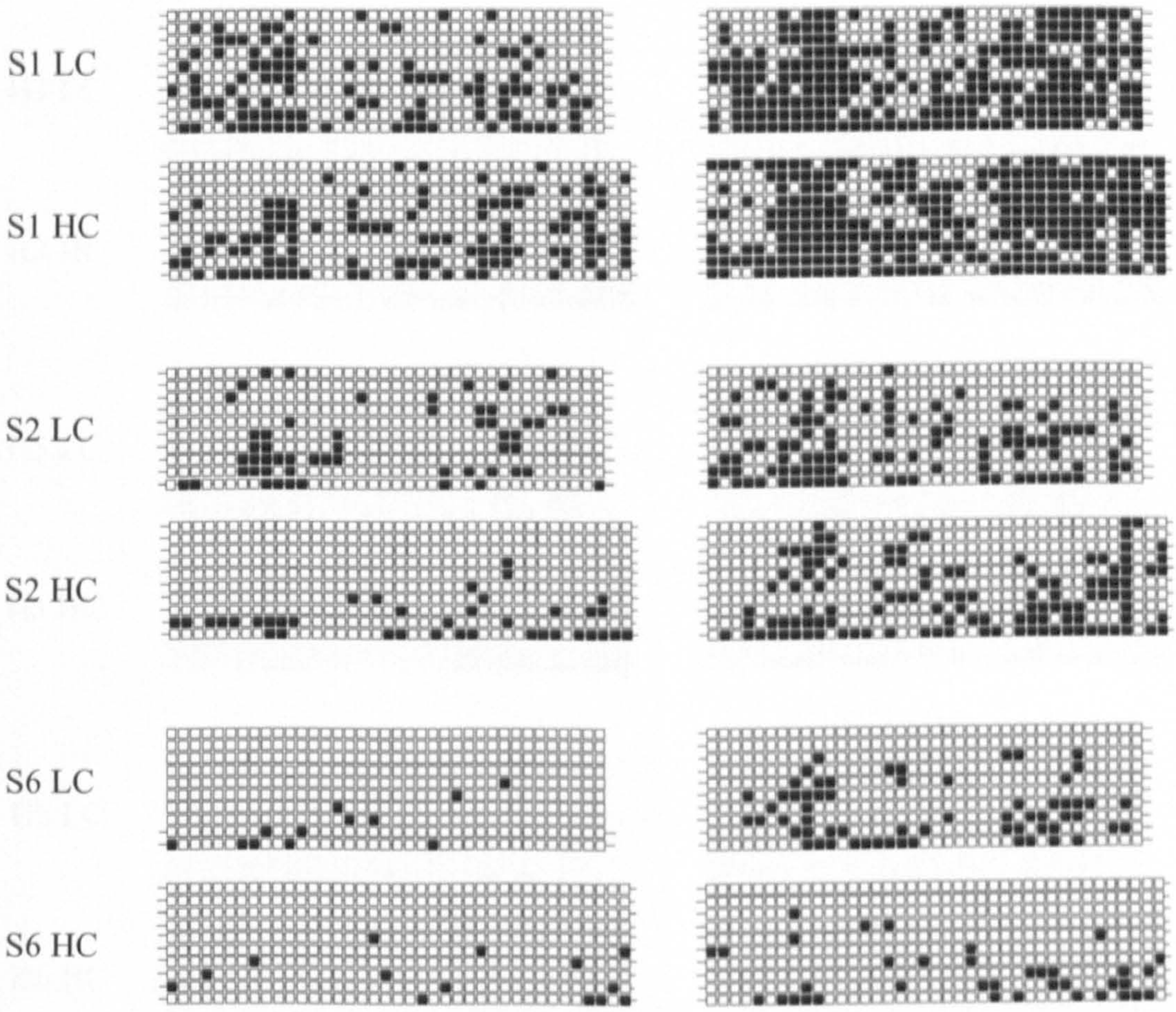
Figure 6.16 Methylation status of CpG dinucleotides within the HC and LC promoter regions in MAR X_S29 clones

Bisulphite sequencing was performed to establish the number of methylated cytosine residues in CpG dinucleotides within the hCMV-MIE promoter. A. Each line represents the result within individual clones of amplified PCR product. White squares = unmethylated CpG dinucleotide, black squares = methylated CpG dinucleotide.

B. Percentage of methylated CpG sites in light and heavy chain promoter region at 0 and 120 generations.

STAR 7 Clones

A.



B.

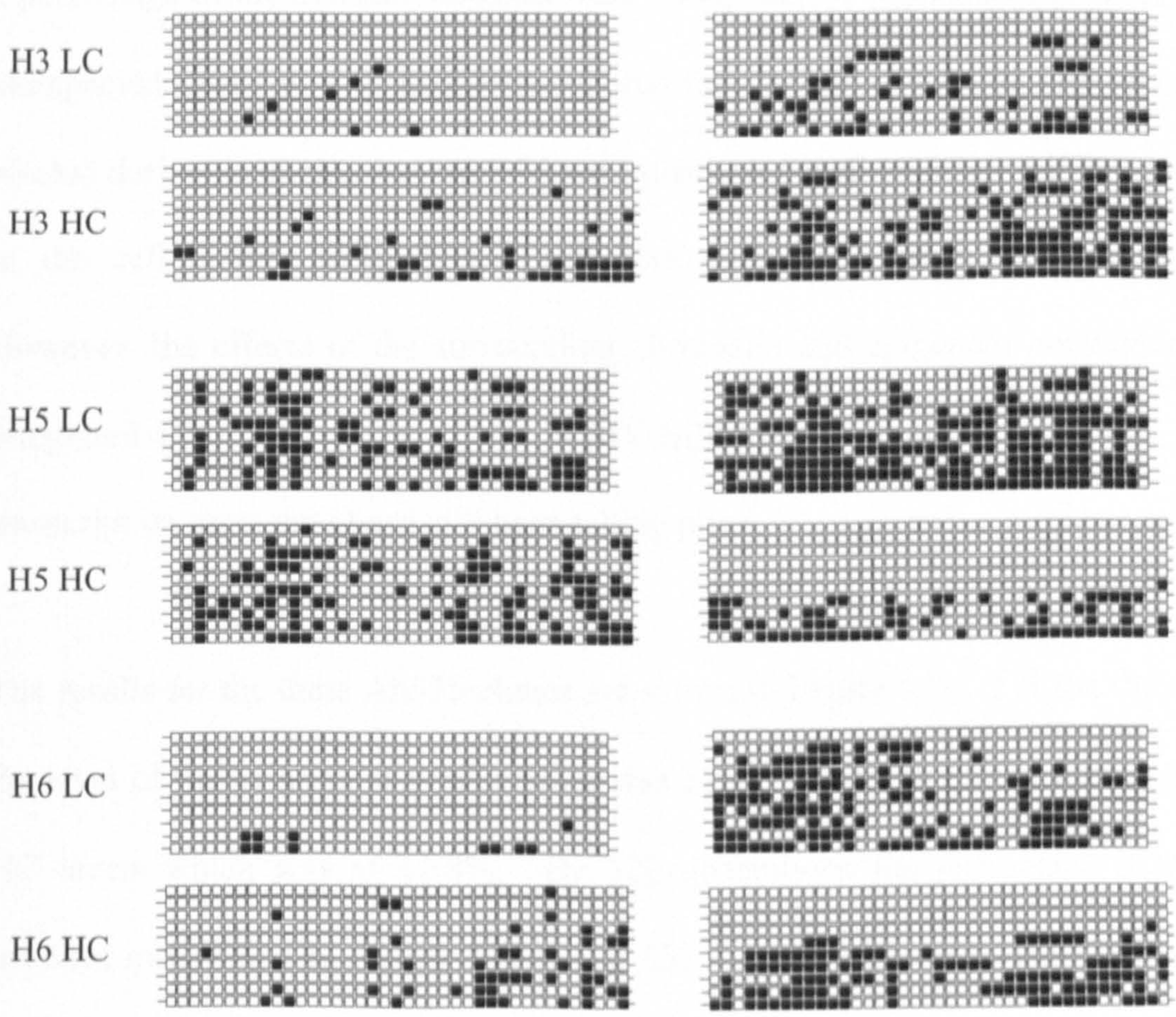
	Light Chain		Heavy Chain	
	0	**120**	**0**	**120**
S1	28.9	63.4	29.2	68.5
S2	14.9	28.1	9.8	30.3
S6	3.2	17.6	3.6	9.5

Figure 6.17 Methylation status of CpG dinucleotides within the HC and LC promoter regions in STAR 7 clones

Bisulphite sequencing was performed to establish the number of methylated cytosine residues in CpG dinucleotides within the hCMV-MIE promoter. A. Each line represents the result within individual clones of amplified PCR product. White squares = unmethylated CpG dinucleotide, black squares = methylated CpG dinucleotide.
B. Percentage of methylated CpG sites in light and heavy chain promoter region at 0 and 120 generations.

cHS4 tandem clones

A.



B.

	Light Chain		Heavy Chain	
	0	**120**	**0**	**120**
H3	1.9	15.1	10	36.9
H5	27.3	50.8	30	38.2
H6	3	35.2	14.3	28.2

Figure 6.18 Methylation status of CpG dinucleotides within the HC and LC promoter regions in cHS4 tandem clones
Bisulphite sequencing was performed to establish the number of methylated cytosine residues in CpG dinucleotides within the hCMV-MIE promoter. A. Each line represents the result within individual clones of amplified PCR product. White squares = unmethylated CpG dinucleotide, black squares = methylated CpG dinucleotide.
B. Percentage of methylated CpG sites in light and heavy chain promoter region at 0 and 120 generations.

Irrespective of whether the chromatin modifying elements were present in the vector, a percentage of the cytosine residues were methylated at 0 generations. This was not unexpected since the cells had been cultured for a period of time (approximately four weeks) during the expansion from 96 well plates to 6 well cultures. MSX was present in the cell culture medium and therefore cells were under selection pressure. However, the effects of the surrounding chromatin and epigenetic regulation of the integrated DNA, especially at the hCMV-MIE promoters driving the HC and LC transcription units may have still been taking place.

The results for the three Ab535 clones are shown in Figure 6.14. For the three clones the level of methylation at 0 generation was below 10% with the exception of 535 7 HC levels which was at 12.8%. By 120 generations the percentage of cytosine residues methylated in clones 535 4 and 535 7 had increased to between 20% and 30%. In contrast, only a small increase in methylation levels were observed in 535 8, with methylation levels at 5% and 10.8% at 120 generations in the LC and HC respectively.

Figure 6.15 shows that with the exception of the HC methylation levels for U8, methylation in the three UCOE clones were all under 2% at 0 generations. By 120 generations the levels had increased. However, overall levels were still below 10%. Methylation levels remained constant over the 120 generations for U7 with only a 0.4% increase in methylation being observed in the LC. In contrast, methylation levels in U8 were higher. Although starting levels in the LC were similar to U2 and U7 at 120 generations the methylation had increased to 14.9%. In the HC the methylation levels were higher with 11.5% of cytosine residues methylated at 0

generations. After 120 generations this had increased to 23.6% which was similar to the levels seen in the Ab535 control clones.

At 0 generations the methylation levels in the MAR X_S29 clones were low (Figure 6.16). Less than 6% of cytosine residues in all three clones were methylated in the HC and LC. After 120 generations there was an increase in methylation levels for all three clones, with the highest increases in methylation being observed in the LC CpG sites with the highest levels of methylation being observed in clone M1 where 22.4% of LC cytosine residues were methylated.

Figure 6.17 shows the methylation status for three STAR 7 clones. In clones S1 and S2 the methylation level was relatively high. The highest percentage was observed in S1 which had 28.9% and 29.2% of the cytosine residues methylated in the LC and HC respectively at 0 generations. After 120 generations the percentage had increased to above 60% in both the LC and HC. The level of methylation in S2 was lower at 0 generations with 14.9% of cytosines methylated in the LC and 9.8% methylated in the HC. However, by 120 generations the percentage of methylated cytosines had risen to 28.1% and 30.3% in the LC and HC. This level of methylation was similar to the amount observed in two of the three Ab535 control clones, 535 4 and 535 7.

Figure 6.18 shows the results of the bisulphite sequencing for the cHS4 tandem clones. For H1 and H6 the levels of methylation in the LC were very low, with only 1.9% and 3% of cytosine residues methylated. In contrast, the levels of methylation in the LC in clone H5 were higher, with 27.3% of cytosines methylated. Levels of methylation were higher in the HC with over 10% of cytosines methylated at 0

generations. Again, H5 has the highest methylation levels with 30% methylated. Samples analysed from gDNA from cells at 120 generations showed that methylation had increased in both the HC and LC in the three clones.

6.2 Discussion

The aim of the work described in this Chapter was to assess whether the presence of chromatin modifying elements in the expression vectors could influence the stability of antibody expression from the clonal stable cell lines.

By culturing the cells long-term in the absence of MSX selective pressure, the expression levels could be quantified during this time period to establish whether antibody expression remained stable. For the majority of clones a decrease in expression was observed irrespective of whether a chromatin modifying element was present in the expression vector or not. Although decreases in expression was observed with all clones, the set of 1.5kb A2UCOE clones showed the lowest decline in expression over the 120 generations. This suggests that the 1.5kb A2UCOE may confer stability, due to the fact that it can protect the expression cassette from repressive epigenetic effects such as the spread of heterochromatin (Antoniou *et al.*, 2003; Williams *et al.*, 2005) and DNA methylation (Zhang *et al.*, 2010). The 1.5kb A2UCOE clones which did show a higher decrease in expression were shown to have unexpected hybridisation signals in the Southern blot analysis. Partial integration of the plasmids may have resulted in the loss of an A2UCOE from the vector. This would have left HC or LC DNA unprotected from the repressive effects of the

surrounding chromatin and susceptibility to DNA methylation resulting in loss of expression.

A subset of clones from each of the five cell lines were analysed further, with specific productivities calculated. The calculation of specific productivities meant that the effect of cell growth in the cultures could be determined. The data showed that there was no conclusive evidence that the presence of the chromatin modifying elements in the expression vectors conferred stability, with one out of the three clones for each of the cell lines exhibiting stability to varying degrees. Within the scope of this study the number of clones being analysed was limited and in order to investigate this further a larger number of clones would need to be analysed. However, it was shown in Chapter 4 that the likelihood and ease of obtaining a high expressing clone is higher when the 1.5kb A2 UCOE is present in the expression vector and therefore it is possible that if a larger number of clones were analysed there would be a larger proportion of clones with stable expression.

Where a significant decrease in expression was observed, in the majority of cases the largest decrease was seen after 40 generations. It would appear that there was a shut down of gene expression during this time and subsequently the loss of gene expression was more gradual.

The loss of stability in antibody expression was accompanied by a reduction in mRNA levels. A reason for the decline in mRNA levels could have been the loss of gene copies from the host cell genome. Southern blot hybridisations performed confirmed that this was not the case, as banding patterns looked identical for all

clones except the HS4 tandem clones. Previous investigations have also shown that a decrease in mRNA levels is not accompanied by a decrease in gene copy numbers (Barnes *et al.*, 2004; Chusainow *et al.*, 2009).

The more intense hybridisation signals observed, in addition to the presence of previously unobserved signals for the HS4 tandem clones at 120 generations was unexpected. The results suggested that there had been an increase in the number of gene copies within the genome. However, as no amplification had been performed on the cell lines there should have been no pressure on the cells to increase gene copies Q-PCR was used to establish that there had not been a large increase in HC and LC gene copy numbers in the clones at 120 generations and therefore the discrepencies may have been an artefact of the Southern blotting procedure.

In order to investigate potential epigenetic mechanisms involved in the gene silencing being observed bisulphite sequencing was performed. This revealed that an increase in methylation in the hCMV-MIE promoter region was evident in a number of the clones where there was a decrease in productivity. The results confirmed the findings of a previous study where it was observed that an increase in DNA methylation contributed to the decrease in antibody expression from stable cell lines (Yang *et al.*, 2010).

In the two A2UCOE clones where specific productivity was above 50% (U2 and U7) at the end of the stability study, DNA methylation levels were low. This was especially particular in clone U7 where levels started at 1% at 0 generations in the HC and LC and levels in the HC remained at 1% in the HC and only increased to 2% in

the LC at 120 generations, indicating that the A2UCOE can prevent the methylation of cytosine residues in HC and LC transcription units.

For the three MAR clones, the levels of DNA methylation were relatively low, with the highest methylation level (22%) at 120 generations observed in the light chain of M1. However, this clone had the highest specific productivity remaining at 120 generations and the remaining two MAR clones, M2 and M7, which had lost over 70% of their specific productivities during the long-term culture had lower levels of DNA methylation. This may indicate that although the MAR X_S29 can protect the HC and LC genes from DNA methylation to some degree, it cannot prevent other mechanisms, from causing the gene silencing being observed.

The highest levels of DNA methylation was observed in one of the STAR 7 cell lines (S1). There have been studies that have suggested that tandem repeats are more susceptible to DNA methylation (McBurney, 2002; Hseih and Fire, 2000). The Q-PCR performed to establish copy number and subsequent Southern blot hybridisations indicated that clone S1 had a high HC and LC copy number (over 70) and the plasmid DNA was integrated in a tandem array. This large amount of repeated DNA sequence in the genome may have resulted in repeat-induced gene silencing and the high level of DNA methylation observed.

Only one of the cHS4 tandem clones conferred any level of stability with the remaining clones exhibiting a large decrease in antibody expression during the long-term culture. This element is an insulator and is said to act as a barrier to chromosomal position effect silencing (Pikaart *et al.*, 1998). It has also been shown

that a tandem copy of the cHS4 core region was sufficient to block the DNA methylation of an *IL-2R* reporter gene that was normally susceptible to chromosomal silencing within 20-40 days in culture (Dickson *et al.*, 2010). In contrast the levels of methylation in the HC and LC promoter regions investigated in this study were higher. The cHS4 element forms part of the erythroid specific chicken β-globin locus control region. As the Dickson *et al.* study was performed in a chicken erythroid cell line (6C2 erythroleukemia), the lack of functionality of this element observed in the work presented in this Thesis may be due to the activity of the cHS4 element being cell type specific. This cell type specificity is supported by another study where it was observed that the cHS4 was only partially effective at insulating the CMV promoter from position effects in CHO cells (Izumi and Gilbert, 1999).

This Chapter has described the work performed to establish the stability of the clones by culturing the cells long-term in the absence of selection pressure and measuring protein and mRNA levels. Southern blotting confirmed that the loss of expression observed in a number of clones was not due to a loss of transgene copies from the genome. Bisulphite sequencing was performed in order to investigate whether the epigenetic regulatory mechanism of DNA methylation was contributing to the gene silencing being observed. Further discussion of these results in relation to the findings presented in the rest of this Thesis is described in Chapter 7.

CHAPTER 7

Discussion

7.1 Discussion

The generation of high expressing, stable cell lines is often a long and laborious process, due to the large number of clones needing to be screened in order to identify a candidate with the desired high and stable expression characteristics. This is, in part, due to transgene site of integration. Expression levels are influenced by the chromatin structure surrounding the integration site. Integration into euchromatin will result in more efficient transcription however, integration into heterochromatin will result in little or no expression. With the majority of mammalian genomes comprising of heterochromatin the probability of isolating a clone where the plasmid has been integrated into a region favourable for high level and stable expression is low (Grewel and Elgin, 2002). In addition, expression can be silenced via a DNA methylation mechanism regardless of whether the transgene has integrated into a chromatin permissive or non-permissive environment (Pikaart *et* al., 1998; Yang *et al.*, 2010). The use of chromatin modifying elements have been shown to increase this probability by shielding the transgene from the effects of the neighbouring chromatin.

When the work described in this Thesis was started there was limited knowledge of how the various chromatin modifying elements described in the literature, namely UCOE, MAR, STAR and the cHS4 insulator element, affected stable antibody production in an industry relevant host cell line. This work constitutes the first in depth side-by-side comparative investigation and characterisation of these elements where their activity and functionality have been studied within the same experimental system.

Although there has been reports of chromatin modifying elements being analysed, the majority of the studies carried out have utilised reporter genes such as GFP and luciferase in order to examine the benefits of these elements (Zahn-Zabal *et al.*, 2001; Kim *et al.*, 2004; Wang *et al.*, 2010; Antoniou *et al.*, 2003; Benton *et al.*, 2002; Williams *et al.*, 2005; Kwaks *et al.*, 2003; Pikaart *et al.*, 1998, Izumi & Gilbert, 1999). Although some studies reported advantages when using chromatin modifying elements in stable antibody expression, experimental parameters between these reports have differed. For example, Girod *et al.* (2005) co-transfected HC and LC genes on separate vectors with the chicken lysozyme MAR element (clys MAR) upstream of the HC expression cassette or even co-transfected on a separate plasmid. Benton *et al.* (2002) describe an expression system where the A2UCOE is located upstream of both transcription units but separate vectors are also used for the expression of the HC and LC genes.

In contrast, UCB expression vectors utilised in stable cell line generation contain both the HC and LC cassettes on the same plasmid vector. It was therefore decided to generate the panel of vectors described in Chapter 3.

Initial experiments to analyse the effects of the chromatin modifying elements on antibody expression were performed in pools of stably transfected cells. The advantage of using stable pools of cells is that it generates recombinant protein quickly, compared to isolating individual clones and the expression level of the pool represents the average of the individual cells within the population. It was shown that the presence of chromatin modifying elements were having an effect on antibody

expression and that different elements performed better than others. The 1.5kb A2UCOE provided the most benefit in comparison to STAR 40 which only marginally increased expression levels compared to the Ab535 control vector. It also revealed that different vector configurations performed best for each of the elements.

In contrast to the studies published previously, (Benton *et al.*, 2002; Williams *et al.*, 2005; Ye *et al.*, 2010) where the UCOE was located immediately upstream of the promoter (equivalent to the 5' LC 5' HC vector used in the work described in this Thesis), results indicated that the optimal vector configuration was when the 1.5kb A2UCOE was at all three locations in the vector (3' HC 5' LC 5' HC). Although UCOEs are thought to function when they are operably linked to a promoter, it would appear that the addition of a further UCOE in the vector resulting in both the LC and HC transcription units being flanked by UCOEs is optimal. This may be due to the fact that it results in the HC and LC transcription units being protected at both the 5' and 3' ends from the negative effects of the surrounding chromatin environment and therefore minimising any position effects.

When comparing the elements side-by-side under GS selection, the incorporation of the 1.5kb A2UCOE in the expression vector resulted in a reduced timeframe for pooled stable recovery as well as increasing expression levels. When cells were cultured in the absence of selection pressure the productivity of the pool of stably transfected cells generated with the construct containing the 1.5kb A2UCOE remained higher than the pooled stable generated using the Ab535 control construct. This was presumably due to the A2UCOE increasing the number of higher expressing clones in the population and therefore, the level and stability of expression within the pool of

cells was greater. Higher expression levels, compared to the Ab535 control were observed in the clonal isolates from the stably transfected cell pools. However, this was not observed when clones were derived using limiting dilution. In this instance although there were a larger number of clones which expressed antibody compared to the control, the highest expressing clones from each vector were comparable.

The results of the initial experiments performed indicated that the MAR X_S29 element was not as potent as the 1.5kb A2UCOE in increasing antibody expression levels in cell pools and clonal stables. A previous report has highlighted the utility of the MAR X_S29 element in increasing recombinant protein expression from CHO cells (Otte *et al.*, 2007). In this investigation the incorporation of MAR X_S29 resulted in a marked increase in expression. However, this was performed with EGFP as a reporter gene, which is the same reporter construct as was used to generate the data presented in the patent application from Selexis (WO 2005/040377) where MAR X_S29 gave the highest increase in GFP levels out of the MAR elements tested. This may indicate that this MAR element does have functionality in CHO cells but when incorporated into antibody expression vectors containing both HC and LC genes its functionality is reduced. In contrast to the data published in the patent from Selexis, it has subsequently been reported that MAR X_S29 has only limited activity in CHO cells compared to the other MAR elements identified in the bioinformatic search. It is reported to perform better in a mesoangioblastic cell line (Mermod, 2007) and another of the MAR elements identified, MAR 1_68, is stated to have better activity in CHO cells (Girod *et al.*, 2007).

The results of the initial experiments to determine the optimal vector configuration for each element showed that STAR 40 had very poor activity in CHO cells. This was in

contrast to the results described in the study where STAR elements were first described which indicated that STAR 40 could increase the number of clones and expression levels of secreted alkaline phosphotase (SEAP) in CHO cells (Kwaks *et al.*, 2003). It was surprising that STAR 7 also failed to show activity, although it is mentioned that STAR 7 is more potent in augmenting gene expression when it is used in combination with another STAR element, STAR 67, which is said to provide a positive, additive effect on antibody expression when placed upstream of a promoter driving transgene expression (Otte *et al.*, 2007; Van Blokland *et al.*, 2007). Therefore, on its own STAR 7 may not be able to enhance transgene expression.

Expression levels for pooled and clonal stable cell lines generated using the cHS4 tandem expression vectors were similar to the Ab535 control and only small increases in the number of expressing clones were observed. The cHS4 element is defined as an insulator and therefore can block activity of a distal enhancer on a promoter, when placed between the two, and can prevent position effects by acting as a 'barrier' (Sun and Elgin, 1999). Therefore, it may not be able to influence the chromatin structure compared to the UCOEs which have been shown to 'open' chromatin and allow expression even within heterochromatic regions (Antoniou *et al.*, 2003; Williams *et al.*, 2005). The data presented in this Thesis also confirms the findings of other studies utilising the cHS4 insulator in CHO cells where it has been shown that the element has limited activity in this cell line (Izumi and Gilbert 1999; Otte *et al.*, 2007). As this insulator cHS4 forms the most 5' distal component of the chicken β-globin locus control region, which may indicate that the functionality of this element is cell type restricted and hence its utility in stable cell line generation in CHO cell lines may be limited.

After the initial experiments were performed to investigate how the chromatin modifying elements affected antibody expression, a thorough characterisation of a number of clones harbouring each of the five test constructs (Ab535 control, 1.5kb A2UCOE, MAR X_S29, STAR 7 and cHS4 tandem) was performed. The focus of this part of the study was to try to understand how these elements might be functioning within the cell. The analysis included genetic characterisation of the clones, using Q-PCR and Southern blotting to determine copy numbers and clonal diversity. Further investigation of the clones was undertaken by culturing the cells in the absence of selection in order to determine whether the presence of the chromatin modifying elements in the expression vectors would influence the stability of the cell lines. DNA methylation levels across the promoter regions in the HC and LC were investigated using bisulphite conversion and DNA sequencing in order to determine whether epigenetic regulatory mechanisms were contributing to any loss of expression. A summary of the findings from these investigations can be found in Table 7.1.

It can be seen that with the exception of two of the STAR 7 clones, S1 and S8, the HC and LC copy numbers were low, with copy numbers below 10 for both HC and LC. The low copy numbers in the 1.5kb A2UCOE cell lines were in agreement with the findings published by Williams *et al.*, (2005) which revealed that A2UCOEs do not increase the number of integrated plasmids. A number of studies have implied that MAR elements increase the number of integrated transgene copies (Girod *et al.*, 2007; Kim *et al.*, 2004; Galbete *et al.* 2009). Although the Q-PCR indicates that HC and LC gene copy number is increased slightly for the MAR X_S29 clones, the results of the Southern blot hybridisations would suggest that copy numbers were similar for the

MAR X_S29 clones compared to the Ab535 control clones. It must be noted that only a small number of clones were analysed due to the limited diversity of the population. Therefore, further investigation of additional clones harbouring the MAR X_S29 element would need to be undertaken to confirm whether this element had an effect on transgene copy number. It has been reported that STAR elements confer copy number dependent expression (Kwaks *et* al., 2003). However, the work described in this Thesis would suggest that this is not the case. Despite the fact that the HC and LC copy numbers appeared to vary, cell clones S1 and S8 had approximately ten-fold more copies than the other clones but expression levels were not any greater (see Table 7.1).

Clone	HC copy number: QPCR	LC copy number: QPCR	Southern blot family member	Expression levels (6 well plates)	% Expression remaining at 120 generations (6 well plates)	% Qp remaining (shake flask cultures)	% Methylaton HC		% Methylation LC	
							0	120	0	120
Ab535 control										
535 1	0.0	0.1	A	0	0					
535 2	0.2	0.3	B	<5	<5					
535 3	0.5	0.4	B	<10	<10					
535 4	**1.1**	**1.1**	**B**	**<10**	**<10**	7	**4**	**25**	**3**	**30**
535 5	0.5	0.6	B	5	<10					
535 6	0.4	0.3	A	0	0					
535 7	**1.4**	**2.0**	**C**	**<15**	**<25**	70	13	28	7	22
535 8	**1.2**	**1.6**	**B**	**<20**	**<35**	11	3	10	2	5
535 9	0.6	0.5	B	5	<5					
535 10	0.6	0.5	B	<15	<10					
1.5kb A2UCOE										
U1	0.6	0.5	A	<30	<85					
U2	**4.6**	**6.4**	**B**	**<30**	**<60**	52	1	9	1	3
U3	1.5	1.8	D	<15	<45					
U4	3.7	6.4	C	<20	<25					
U5	0.5	0.6	A	<40	<70					
U6	1.3	1.6	A	35	60					
U7	**3.7**	**6.7**	**A**	**40**	**<80**	69	1	1	1	2
U8	**1.2**	**3.4**	**D**	**<15**	**<30**	30	12	24	2	15
U9	1.2	1.2	D	<25	<65					
U10	1.2	1.0	A	35	80					
MAR X_S29										
M1	**3.8**	**3.6**	**A**	**15**	**25**	63	6	12	3	22
M2	**2.9**	**3.6**	**B**	**<10**	**<100**	21	2	11	2	13
M3	2.9	2.4	A	<20	<25					
M4	3.1	4.6	A	<20	<15					
M5	1.2	1.6	B	<5	<10					
M6	3.3	3.3	B	<5	<30					
M7	**2.0**	**1.8**	**C**	**<20**	**15**	29	1	8	1	16
M8	2.8	3.0	B	<10	<10					
M9	2.2	2.4	B	<5	20					
M10	1.2	2.8	B	<5	<60					
STAR 7										
S1	**71.0**	**77.5**	**A**	<5	**<20**	19	29	69	29	63
S2	**4.9**	**5.4**	**B**	<5	**<20**	3	10	30	15	28
S3	7.3	6.9	C	<5	<30					
S4	0.0	0.0	D	0	0					
S5	1.6	2.3	B	<5	<30					
S6	**7.5**	**9.2**	**E**	**<5**	**65**	52	4	18	3	10
S7	4.0	4.3	B	<5	10					
S8	97.3	122.6	A	<5	<15					
S9	6.4	9.1	B	<5	<15					
S10	8.1	9.0	C	<5	30					
cHS4 Tandem										
H1	0.8	0.5	A	5	<35					
H2	0.5	0.9	B	<10	<15					
H3	**1.5**	**1.4**	**A**	**<10**	**<20**	7	10	37	2	15
H4	5.4	2.1	C	5	<50					
H5	**2.1**	**0.9**	**C**	**<5**	**<65**	23	30	38	27	51
H6	**2.8**	**2.3**	**D**	<15	**<25**	80	14	28	3	35
H7	1.2	1.4	A	<5	<40					
H8	1.0	1.0	A	<5	20					
H9	7.7	4.6	B	<15	<20					
H10	0.7	1.3	B	15	5					

Table 7.1 Summary of findings

Summary of the data generated for the ten clones analysed harbouring each of the test contrusts. Results showing copy number as determined by Q-PCR, Southern blot family, expression levels from overgrown cultures in 6 well plates, expression (%) remaining after 120 generations, percentage of Q_p remaining at 120 generations and methylation status of HC and LC at 0 and 120 generations. Clones in different colours represent the different families established by Southern blot hybridisations.

The results of the Southern blot analyses were also very insightful. Clones had to be derived from the stably transfected GS selected pools of cells due to the lack of cell viability in a large number of clones derived using limiting dilution. The results highlighted that the diversity of cell populations in these stable cell pools was much lower than expected. Table 7.1 shows that none of the five cell lines analysed contained ten individual clones with different integration sites as a number of clones in each test group shared the same banding patterns. This suggested that a number of the clones investigated were likely to be daughter progeny from the same original transfected cell. However, to confirm whether this was indeed the case, sequence analysis of the site of integration for each clone would need to be performed. The results of the Southern blot hybridisations indicated that the time at which clones are isolated from stable cell pools is important in order to isolate a diverse population of clones. The data presented here suggests that by the time the cell viability of the population had recovered to greater than 90%, the clonal diversity had already decreased with the majority of the cells making up the population being related rather than independent clones. This may also provide an explanation for the results observed in Chapter 4, which showed that the 1.5kb A2UCOE containing clones derived from the GS stably transfected pool expressed at a higher level than the clones derived from the Ab535 control. At the time of deriving single colonies, if only a few high producing clones were present in the Ab535 control stable cell pool as a result of these cells being out grown by faster growing lower producing cells, the likelihood of isolating a high producing clone would be low.

Culturing the cells in the absence of selection pressure allowed the stability of the clones to be investigated. In the majority of clones expression levels decreased over

the 120 generations. In general, the clones which belonged to the same family as determined by having the same banding pattern in the Southern blot analysis, exhibited a similar expression profile over the 120 generations, with comparable starting levels and loss of expression during the culture period. It can also be seen in Table 7.1 that for the majority of clones a decrease in expression was observed, which was irrespective of whether a chromatin modifying element was present. For a subset of clones specific productivities were calculated so that the effect of cell growth on expression could be determined. The combination of the data generated in 6 well plates and the specific productivity analysis indicated that there was no conclusive evidence that the presence of the chromatin modifying elements in the expression vectors conferred stability. The data does hint at the 1.5kb A2UCOE conferring stability as the set of clones harbouring this test construct showed the lowest decrease in expression over the 120 generations. The exception to this was the clones that had the multiple hybridisation signals in the Southern blot analysis (U3, U8 and U9) which indicated that regions of the integrated plasmid molecules had been partially deleted and may have lost one or more 1.5kb A2UCOEs from the vector. The lack of stability across the cHS4 tandem clones may again point to the cell type specificity of this element. Previous studies have reported the utility of this insulator in preventing position effects but these have been in non-CHO cell lines, with the majority being erythroid in nature (Villemure *et al.*, 2001; Walters *et al.*, 1999; Dickson *et al.*, 2010).

Southern blot analysis was performed at the end of the stability study in order to determine whether the loss of expression was due to a loss of transgenes from the genome. For all cell lines except for clones harbouring the cHS4 tandem construct no differences were observed, which confirmed that transgene copies had not been lost

from the genome during the long-term culture in the absence of selection pressure. In several of the cHS4 tandem clones additional bands were observed and hybridisation signals were more intense indicating that there may have been an increase in the number of transgenes in the genome. This result may have been expected if gene amplification had been performed on the cell lines. However, as the cells had been cultured in the absence of selection, there should have been no pressure to increase gene copies. In an attempt to confirm whether transgene copies had increased in the cHS4 clones, Q-PCR was performed using the gDNA extracted from cells at the end of the stability study. This showed that there had not been a large increase in transgene copies during the long-term culture, and therefore the discrepencies may have been an artefact of the Southern blotting procedure.

In order to determine whether gene expression was being regulated by epigenetic mechanisms, bisulphite conversion followed by DNA sequencing of the hCMV-MIE promoter regions that drive expression of the HC and LC genes was performed. As summarised in Table 7.1, in a number of clones that showed a decrease in expression an increase in methylation across the hCMV-MIE promoter regions was observed. This suggests that for a number of clones the methylation of cytosine residues in the promoter region was contributing at least in part to the observed loss of antibody productivity. This was especially evident in clone S1 transfected with the STAR 7 construct, which had high HC and LC gene copy numbers. It is known that tandem repeats of integrated transgenes are more susceptible to DNA methylation (McBurney, 2002; Hseih and Fire, 2000). It would appear from the high levels of DNA methylation observed in these clones that the large quantity of repeated DNA sequence in the genome had resulted in these sequences becoming methylated and

therefore silenced. For the other clones where the decrease in expression was accompanied by an increase in levels of DNA methylation, the starting levels and increase in methylation were less than that observed in clone S1. The findings presented here are consistent with another report which indicated that the loss of productivity from a number of antibody expressing stable cell lines was a result of DNA methylation (Yang *et al.*, 2010). For a number of clones the decrease in expression was not accompanied by a large increase in DNA methylation. For example, clone 535 4 and 535 8 both exhibited a decrease in specific productivity over the long-term culture period. However, although DNA methylation levels were similar at the start of the study, a larger increase in methylation was observed in 535 4 whereas levels in 535 8 remained relatively unchanged. DNA methylation levels of two of the MAR construct containing clones (M2 and M7) were also low, even though a large decrease in specific productivity was observed. This suggests that other epigenetic regulatory mechanisms may be involved which resulted in the gene silencing observed. This has been demonstrated in a study by Mutskov *et al.* (2003) where they revealed that the silencing of transgene expression was due to a decrease in histone acetylation and methylation of histone tails rather than via DNA methylation. A second potential explanation for the lower DNA methylation levels observed is that although the bisulphite sequencing gives a representative view of the DNA methylation status, the sample size analysed was small and with only ten sequences analysed it represents only a very small proportion of cells. It may be that the ten sequences analysed do not show a true representation of the methylation status across the promoter regions and levels may in fact be higher in these clones.

The lowest DNA methylation levels out of all clones were observed in the 1.5kb A2UCOE clone, U7, where only 1% and 2% of the cytosine residues were methylated in the HC and LC respectively at 120 generations. Methylation levels were also low in clone U2, which had a specific productivity above 50% at the end of the stability study. The remaining clone, U8, lost productivity over the 120 generations and was shown to have higher levels of methylation. The Southern blot analysis revealed the presence of multiple bands in clone U8, which indicates integration of partially deleted plasmid molecules, or integration of multiple copies in an array that was not as head-to-head or head-to-tail. Integration of deleted plasmids may have resulted in the loss of an A2UCOE from the vector which would have left the HC or LC cassettes unprotected from the repressive effects of the surrounding chromatin and therefore susceptible to DNA methylation. A recent report has also demonstrated the ability of the 1.5kb A2UCOE to prevent DNA methylation of an integrated transgene (Zhang *et al.*, 2010). In this study, it was found that the stability of transgene expression in murine embryonic carcinoma P19 cells from the A2UCOE within self-inactivating lentiviral vectors (SIN-LVs) was due to the transgene being resistant to DNA methylation. Epigenetic analysis of the *HNRPA2B1-CBX3* locus has found that a large region of the DNA is free from methylation beyond the boundaries of the CpG island spanning the divergent promoters of these genes, and histone modifications associated with active transcription are present (Lindahl Allen and Antoniou, 2007). This has led to the theory that the functionality of the UCOEs is due to the reproduction of these epigenetic marks at the site of transgene integration which will then prevent DNA methylation and heterochromatin formation and prevent transcriptional silencing (Zhang *et al.*, 2010).

The analysis of transgene DNA methylation levels showed that DNA methylation is contributing to the loss in productivity in a number of clones. It also showed that levels of methylation varied across the clones at the start of the stability study. Clones such as (STAR) S1 and (cHS4) H5 whose specific productivity decreased over the 120 generations, had a relatively high level of methylation. For a successful manufacturing cell line the expression levels must remain stable over a long period of time. Currently, in order to establish whether transgene expression in a line is stable, cells must be cultured long-term in the absence of selection to determine whether productivity remains constant over this period. This prolongs the time required to generate the manufacturing cell line. Ideally, it would be advantageous to have a method that could predict the stability of expression in a cell line in the early stages of its generation. As it has been shown that levels of methylation vary between clones, quantification of the levels of DNA methylation early in a cell line development process may allow the potential stability to be predicted. Bisulphite sequencing is quite a labour intensive method and may not be conducive for analysing large numbers of clones. However, bisulphite conversion followed by a quantitative methyl specific PCR (Q-MSP) assay which was sensitive to levels of methylated DNA could be utilised. Q-MSP combines methyl specific PCR (MSP) (Herman *et al.*, 1996) with Q-PCR to allow high-throughput quantitative analysis of the levels of methylation in the region studied. Oligonucleotides which can distinguish between methylated and unmethylated cytosine residues in the bisulphite converted DNA are used, which allow differential amplification of the target DNA depending on levels of methylation.

The data presented in this Thesis has shown that the addition of certain chromatin modifying elements to expression vectors is beneficial for stable cell line generation,

and they can facilitate the isolation of high expressing stable clones. The importance of being able to improve the predictability of stable cell line generation and reduce timelines and workload is an ongoing aim within the industry and as such, research has focussed on ways of achieving this. The work reported in this Thesis has concentrated on one such solution; however, other approaches to solve this problem have been reported. Studies utilising endogenous CHO DNA sequences to enhance transgene expression have also been reported (Running Deer and Allison, 2004). In this study the addition of the 5' and 3' regions of the highly expressed housekeeping gene *elongation factor 1 alpha* (*EF1α*), which flanked a reporter gene in the expression vector were shown to increase transgene expression. Average expression levels from cell pools and stable clones for a number of different reporter genes were increased between 6 and 35-fold compared to expression levels obtained using commercial vectors which utilised the CMV or *EF1α* promoters.

Other methods to improve the probability of isolating clones with high and stable levels of expression are to utilise site specific integration (SSI) to introduce genes into a pre-defined locus within the genome, which have been previously identified as a favourable site. Studies have highlighted the use of Zinc Finger Nucleases (ZFNs) (Moehle *et al.*, 2007), Meganucleases (Cabaniols *et al.*, 2010) and also AttSite® Recombinase Technology (Campbell *et al.*, 2010). ZFNs are comprised of two functional domains, one for binding to DNA, and another which comprises a nuclease domain of the *Fok*I restriction enzyme, which performs the DNA cleavage. This generates a double stranded break at a predetermined site in genome, which can greatly augment the efficiency of integration at this location by homologous recombination in the host cell genome. Meganucleases are similar to ZFNs in that

they bind to a specific DNA recognition site and induce a double strand break in the DNA. Their recognition site can range from 12-30bp, which makes them unique and highly site specific. AttSite® Recombinase technology preferentially integrates plasmid DNA into the genome at predefined loci by utilising gene-targeting serine recombinase enzymes to catalyse the stable and irreversible insertion of DNA at specific *attB* sites in the host genome. These methods for SSI have only recently been described and there are limited reports of their utility in the production of recombinant proteins. A report has highlighted the advantages of using a natural yeast meganuclease, I-*Sce*I, to induce gene insertion in CHO cells. It was shown that expression levels of clones expressing either β-galactosidase or human CD4 were reproducible and these levels were stable for over four months (Cabaniols and Paques, 2008). The utility of the AttSite® Recombinase technology has also been reported recently where it was shown that clones expressing a therapeutic protein derived using this technology had uniform, high level expression compared to clones generated by random integration of the transgene (Campbell *et al.*, 2010).

A related technology forgoes the need to integrate a transgene into the native chromatin environment all together. The Artifical Chromosome Expression (ACE) technology utilises pre-engineered artificial chromosomes which reside in a production cell line and have multiple recombination acceptor sites allowing the targeted insertion of multiple copies of DNA sequences into these sites. The artificial chromosomes contain fully functional centromeres and telomeres, and as a result are said to be as mitotically stable as the host chromosomes. The ability of the ACE technology to generate high expressing stable CHO cell lines has been described. Kennard *et al.* (2009) have reported that cell lines expressing a human monoclonal

IgG1 antibody at over 1g/l in batch culture could be generated using this technology. Stability analysis showed that expression remained relatively stable in the presence of selection pressure. However, when this was removed from the media the same cell lines were considerably more unstable, losing up to 70% in expression. This system incorporates the cHS4 insulator within the expression vector, flanking the HC and LC expression cassette. Therefore, although in this context it is used in combination with the artificial chromosomes, it would appear to be another example of the cHS4 insulator having limited functionality in CHO cells.

7.2 Future Work

In the work described in this Thesis epigenetic regulation, in the form of DNA methylation, and its effects on antibody expression has been investigated. The results show that increasing levels of DNA methylation contribute to HC and LC gene silencing and subsequent loss of expression from the clones. DNA methylation is only one of the repressive epigenetic marks and histone modifications such as deacetylation and methylation of H3K9 and especially H3K27me3 also play a role in the silencing of transgenes (Mutskov and Felsenfeld, 2004) and as described above, may be causing gene silencing in several of the clones analysed. Within the scope of this study the effects of histone modifications on the regulation of transgene expression have not been analysed. However, further investigation of this would provide a clearer understanding of the mechanisms contributing to transgene silencing within the cell. Studies have shown that the addition of iHDACs such as sodium butyrate (Mimura *et* al, 2001; Palmero *et* al., 1991; Jiang and Sharfstein, 2008) can have a beneficial effect on antibody expression levels due to the inhibition of removal

of acetyl moieties from histone tails. Assessing the effect of the addition of iHDACs to the culture medium of clones which have exhibited a loss of expression, would establish whether histone deacetylation was contributing to gene silencing. In addition, analysis of the PTMs on histone tails in the vicinity of the HC and LC would provide an insight into what, if any, effects histone modifications were having on the regulation of antibody expression. Chromatin Immunoprecipitation (ChIP) could be used to analyse specific histone modifications (such as H3K4 acetylation as an example of an active epigenetic mark and H3K27 trimethylation as an example of a repressive epigenetic mark) to establish if there is a correlation between these marks and levels of expression from the clones. Although ChIP analysis was attempted, this proved to be unsuccessful, with no enrichment being observed even in positive control samples. This may have been due to the specificity of the antibodies. No anti-hamster specific antibodies could be sourced. Further investigation into obtaining suitable antibodies and optimisation of the protocol for CHO cells would need to be undertaken in order to determine the histone modifications taking place in the cell lines.

Having shown that the presence of chromatin modifying elements is beneficial to antibody expression it would be advantageous to identify a novel element which could be utilised in the UCB mammalian expression vectors without the need to licence an existing chromatin modifying element. One possible way in which a novel element could be identified is to determine the site of trangene integration in a high producing cell line. This could be achieved by methods such as inverse PCR (iPCR) or ligation mediated PCR (LM-PCR). These methods would allow the sequence of the gDNA surrounding the site of integration to be determined. The region of DNA identified

could then be incorporated into the expression vector and tested for its ability to confer high and stable expression in stably transfected cells.

7.3 Conclusions

The work described in this Thesis has highlighted the utility of chromatin modifying elements in the generation of stable cell lines. It is clear that the presence of the 1.5kb A2UCOE at three locations (3'HC 5'LC 5'HC) within the antibody expression vector conferred a clear advantage over the Ab535 control in the generation of both pools and clonal stable cell lines. It has been shown that the other chromatin modifying elements under investigation have only limited functionality in CHO cells and are not beneficial in facilitating stable cell line generation for the expression of antibodies from CHO cells. It has also been shown that the chromatin modifying elements only have a modest effect on expression stability when cells were grown in the absence of selection pressure, with the productivity decreasing in all clones. However, across the clones analysed the lowest decreases were observed in the 1.5kb A2UCOE containing clones. The decrease in antibody expression was accompanied by a decrease in mRNA levels which was not caused by a loss of either HC or LC gene copies from the genome. Analysis of DNA methylation patterns across the hCMV-MIE promoter regions demonstrated that this epigenetic regulatory mechanism was involved in the silencing of a number of clones which exhibited a decrease in productivity.

CHAPTER 8

References

Chapter 8: References

Aapola U., Lyle R., Krohn K., Antonarakis S.E., & Peterson P. (2001) Isolation and initial characterization of the mouse Dnmt3l gene. *Cytogenet.Cell Genet.* 92, 122-126.

Allfrey V.G., FaulknerR R., & Mirsky A.E. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc.Natl.Acad.Sci.U.S.A* 51, 786-794.

Antequera F. & Bird A. (1993) Number of CpG islands and genes in human and mouse. *Proc.Natl.Acad.Sci.U.S.A* 90, 11995-11999.

Antequera F. (2003) Structure, function and evolution of CpG island promoters. *Cell Mol Life Sci* 60, 1647-1658.

Antoniou M., Harland L., Mustoe T., Williams S., Holdstock J., Yague E., Mulcahy T., Griffiths M., Edwards S., Ioannou P.A., Mountain A., & Crombie R. (2003) Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. *Genomics* 82, 269-279.

Arents G., Burlingame R.W., Wang B.C., Love W.E., & Moudrianakis E.N. (1991) The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc.Natl.Acad.Sci.U.S.A* 88, 10148-10152.

Bannister A.J. & Kouzarides T. (1996) The CBP co-activator is a histone acetyltransferase. *Nature* 384, 641-643.

Bannister A.J. & Kouzarides T. (2005) Reversing histone methylation. *Nature* 436, 1103-1106.

Barnes,L.M., Bentley,C.M., and Dickson,A.J. (2001). Characterization of the stability of recombinant protein production in the GS-NS0 expression system. Biotechnol. Bioeng. *73*, 261-270

Barnes L.M., Bentley C.M., & Dickson A.J. (2003) Stability of protein production from recombinant mammalian cells. *Biotechnol Bioeng.* 81, 631-639.

Barnes L.M., Bentley C.M., & Dickson A.J. (2004) Molecular definition of predictive indicators of stable protein expression in recombinant NS0 myeloma cells. *Biotechnol.Bioeng.* 85, 115-121.

Barski A., Cuddapah S., Cui K., Roh T.Y., Schones D.E., Wang Z., Wei G., Chepelev I., & Zhao K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823-837.

Bassett A., Cooper S., Wu C., & Travers A. (2009) The folding and unfolding of eukaryotic chromatin. *Curr.Opin.Genet.Dev.* 19, 159-165.

Bauer U.M., Daujat S., Nielsen S.J., Nightingale K., & Kouzarides T. (2002) Methylation at arginine 17 of histone H3 is linked to gene activation. *EMBO Rep.* 3, 39-44.

Baylin S.B. (2005) DNA methylation and gene silencing in cancer. *Nat.Clin.Pract.Oncol.* 2 Suppl 1, S4-11.

Beard C., Li E., & Jaenisch R. (1995) Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes and Development* 9, 2325-2334.

Bebbington C.R., Renner G., Thomson S., King D., Abrams D., & Yarranton G.T. (1992) High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Biotechnology (N.Y.)* 10, 169-175.

Beck A., Wagner-Rousset E., Bussat M.C., Lokteff M., Klinguer-Hamour C., Haeuw J.F., Goetsch L., Wurch T., Van D.A., & Corvaia N. (2008) Trends in glycosylation, glycoanalysis and glycoengineering of therapeutic antibodies and Fc-fusion proteins. *Curr.Pharm.Biotechnol.* 9, 482-501.

Bell A.C. & Felsenfeld G. (1999) Stopped at the border: boundaries and insulators. *Curr.Opin.Genet.Dev.* 9, 191-198.

Bell A.C., West A.G., & Felsenfeld G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* 291, 447-450.

Benton T, Tim Chen, Michele McEntee, Brian Fox, David King, Robert Crombie, Thomas C.Thomas, & Christopher Bebbington (2002) The Use of UCOE vectors in combination with a preadapted serum free, suspension cell line allows for rapid prouction of large quantities of protein. *Cytotechnology.*

Berger S.L., Kouzarides T., Shiekhattar R., & Shilatifard A. (2009) An operational definition of epigenetics. *Genes and Development* 23, 781-783.

Bernstein B.E., Meissner A., & Lander E.S. (2007) The Mammalian Epigenome. *Cell* 128, 669-681.

Bestor T., Laudano A., Mattaliano R., & Ingram V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J.Mol.Biol.* 203, 971-983.

Bestor T.H. (2000) The DNA methyltransferases of mammals. *Human Molecular Genetics* 9, 2395-2402.

Birch J.R. & Racher A.J. (2006) Antibody production. *Adv.Drug Deliv.Rev.* 58, 671-685.

Bird A., Taggart M., Frommer M., Miller O.J., & Macleod D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91-99.

Bird A.P. & Wolffe A.P. (1999) Methylation-Induced Repression-- Belts, Braces, and Chromatin. *Cell* 99, 451-454.

Bode J., Stengert-Iber M., Kay V., Schlake T., & Dietz-Pfeilstetter A. (1996) Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. *Crit Rev.Eukaryot.Gene Expr.* 6, 115-138.

Bode J., Benham C., Knopp A., & Mielke C. (2000) Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). *Crit Rev.Eukaryot.Gene Expr.* 10, 73-90.

Boshart M., Weber F., Jahn G., Dorsch-Hasler K., Fleckenstein B., & Schaffner W. (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41, 521-530.

Brandeis M., Frank D., Keshet I., Siegfried Z., Mendelsohn M., Nemes A., Temper V., Razin A., & Cedar H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature* 371, 435-438.

Brown M.E., Renner G., Field R.P., & Hassell T. (1992) Process development for the production of recombinant antibodies using the glutamine synthetase (GS) system. *Cytotechnology* 9, 231-236.

Burgers W.A., Fuks F., & Kouzarides T. (2002) DNA methyltransferases get connected to chromatin. *Trends Genet* 18, 275-277.

Cabaniols J.P., Ouvry C., Lamamy V., Fery I., Craplet M.L., Moulharat N., Guenin S.P., Bedut S., Nosjean O., Ferry G., Devavry S., Jacqmarcq C., Lebuhotel C., Mathis L., Delenda C., Boutin J.A., Duchateau P., Coge F., & Paques F. (2010) Meganuclease-Driven Targeted Integration in CHO-K1 Cells for the Fast Generation of HTS-Compatible Cell-Based Assays. *J.Biomol.Screen.* 15, 956-967.

Cabaniols,J.P. and Paques,F. (2008). Robust cell line development using meganucleases. *Methods Mol. Biol.* 435, 31-45.

Campanero M.R., Armstrong M.I., & Flemington E.K. (2000) CpG methylation as a mechanism for the regulation of E2F activity. *Proc.Natl.Acad.Sci.U.S.A* 97, 6481-6486.

Campbell M., Corisdeo S., McGee C., & Kraichely D. (2010) Utilization of site-specific recombination for generating therapeutic protein producing cell lines. *Mol.Biotechnol.* 45, 199-202.

Campos E.I. & Reinberg D. (2009) Histones: annotating chromatin. *Annu.Rev.Genet.* 43, 559-599.

Cao R., Wang L., Wang H., Xia L., Erdjument-Bromage H., Tempst P., Jones R.S., & Zhang Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039-1043.

Chen C. & Chasin L.A. (1998) Cointegration of DNA molecules introduced into mammalian cells by electroporation. *Somat.Cell Mol.Genet.* 24, 249-256.

Cheung P., Allis C.D., & Sassone-Corsi P. (2000) Signaling to chromatin through histone modifications. *Cell* 103, 263-271.

Choi K.H., Basma H., Singh J., & Cheng P.W. (2005) Activation of CMV promoter-controlled glycosyltransferase and beta -galactosidase glycogenes by butyrate, tricostatin A, and 5-aza-2'-deoxycytidine. *Glycoconj.J.* 22, 63-69.

Chuang L.S., Ian H.I., Koh T.W., Ng H.H., Xu G., & Li B.F. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 277, 1996-2000.

Chung J.H., Whiteley M., & Felsenfeld G. (1993) A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* 74, 505-514.

Chung J.H., Bell A.C., & Felsenfeld G. (1997) Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A* 94, 575-580.

Chusainow J., Yang Y.S., Yeo J.H., Toh P.C., Asvadi P., Wong N.S., & Yap M.G. (2009) A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnol.Bioeng.* 102, 1182-1196.

Clouaire T. & Stancheva I. (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol.Life Sci.* 65, 1509-1522.

Davie J.R. (2003) Inhibition of histone deacetylase activity by butyrate. *J.Nutr.* 133, 2485S-2493S.

Dennis K., Fan T., Geiman T., Yan Q., & Muegge K. (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes and Development* 15, 2940-2944.

Derouazi M., Martinet D., Besuchet S.N., Flaction R., Wicht M., Bertschinger M., Hacker D.L., Beckmann J.S., & Wurm F.M. (2006) Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem.Biophys.Res.Commun.* 340, 1069-1077.

Dickson J., Gowher H., Strogantsev R., Gaszner M., Hair A., Felsenfeld G., & West A.G. (2010) VEZF1 elements mediate protection from DNA methylation. *PLoS.Genet.* 6, e1000804.

Dokmanovic M., Clarke C., & Marks P.A. (2007) Histone deacetylase inhibitors: overview and perspectives. *Mol.Cancer Res.* 5, 981-989.

Edmondson D.G., Smith M.M., & Roth S.Y. (1996) Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4. *Genes and Development* 10, 1247-1259.

Eisen J.A., Sweder K.S., & Hanawalt P.C. (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* 23, 2715-2723.

Elgin S.C.R. & Grewal S.I.S. (2003) Heterochromatin: silence is golden. *Current Biology* 13, R895-R898.

Felsenfeld G. (1993) Chromatin structure and the expression of globin-encoding genes. *Gene* 135, 119-124.

Filippova G.N., Fagerlie S., Klenova E.M., Myers C., Dehner Y., Goodwin G., Neiman P.E., Collins S.J., & Lobanenkov V.V. (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol.Cell Biol.* 16, 2802-2813.

Forrest-Owen,W., Daramola,L., Hatton,D., and Field,R. (2005) Evaluation of the Stability of NS0 Cell Lines Expressing Recombinant Human IgG. ESACT Proceedings, *Animal Cell Technology*. 2, 513-515.

Fraga M.F., Ballestar E., Montoya G., Taysavang P., Wade P.A., & Esteller M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.* 31, 1765-1774.

Frank D., Keshet I., Shani M., Levine A., Razin A., & Cedar H. (1991) Demethylation of CpG islands in embryonic cells. *Nature* 351, 239-241.

Fritsch O., Benvenuto G., Bowler C., Molinier J., & Hohn B. (2004) The INO80 protein controls homologous recombination in Arabidopsis thaliana. *Mol.Cell* 16, 479-485.

Fuks F., Hurd P.J., Wolf D., Nan X., Bird A.P., & Kouzarides T. (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J.Biol.Chem.* 278, 4035-4040.

Galbete,J.L., Bucceta,M., and Mermod,N. (2009). MAR elements regulate the probability of epigenetic switching between active and inactive gene expression. *Mol. Biosyst.* 5, 143-150.

Gardiner-Garden M. & Frommer M. (1987) CpG islands in vertebrate genomes. *J.Mol.Biol.* 196, 261-282.

Geyer P.K. & Corces V.G. (1992) DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes and Development* 6, 1865-1873.

Geyer P.K. & Parnell T.J (2001) Chromosomes: Higher Order Organization. *Encyclopedia of Life Sciences* http://www.els.net/ [DOI:10.1038/npg.els.0001157].

Girod P.A., Zahn-Zabal M., & Mermod N. (2005) Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines. *Biotechnol Bioeng.* 91, 1-11.

Girod P.A., Nguyen D.Q., Calabrese D., Puttini S., Grandjean M., Martinet D., Regamey A., Saugy D., Beckmann J.S., Bucher P., & Mermod N. (2007) Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat Meth* 4, 747-753.

Gorman C. & Bullock C. (2000) Site-specific gene targeting for gene expression in eukaryotes. *Curr.Opin.Biotechnol.* 11, 455-460.

Grewal S.I. & Elgin S.C. (2002) Heterochromatin: new possibilities for the inheritance of structure. *Curr.Opin.Genet.Dev.* 12, 178-187.

Harland L., Crombie R., Anson S., deBoer J., Ioannou P.A., & Antoniou M. (2002) Transcriptional regulation of the human TATA binding protein gene. *Genomics* 79, 479-482.

Hart C.M. & Laemmli U.K. (1998) Facilitation of chromatin dynamics by SARs. *Curr Opin Genet Dev* 8, 519-525.

Hendrich B. & Bird A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol.Cell Biol.* 18, 6538-6547.

Hendrich B., Hardeland U., Ng H.H., Jiricny J., & Bird A. (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401, 301-304.

Hendrich B., Guy J., Ramsahoye B., Wilson V.A., & Bird A. (2001) Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes and Development* 15, 710-723.

Hendzel M.J., Wei Y., Mancini M.A., Van H.A., Ranalli T., Brinkley B.R., Bazett-Jones D.P., & Allis C.D. (1997) Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* 106, 348-360.

Herman J.G., Graff J.R., Myohanen S., Nelkin B.D., & Baylin S.B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc.Natl.Acad.Sci.U.S.A* 93, 9821-9826.

Hsieh J. & Fire A. (2000) Recognition and silencing of repeated DNA. *Annu.Rev.Genet.* 34, 187-204.

Huang S., Li X., Yusufzai T.M., Qiu Y., & Felsenfeld G. (2007) USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. *Mol.Cell Biol.* 27, 7991-8002.

Iguchi-Ariga S.M. & Schaffner W. (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes and Development* 3, 612-619.

Illingworth R.S. & Bird A.P. (2009) CpG islands--'a rough guide'. *FEBS Lett.* 583, 1713-1720.

Izumi M. & Gilbert D.M. (1999) Homogeneous tetracycline-regulatable gene expression in mammalian fibroblasts. *J.Cell Biochem.* 76, 280-289.

Jackson J.P., Lindroth A.M., Cao X., & Jacobsen S.E. (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416, 556-560.

Jayapal K.R, Wlaschin K.F, Hu W-S, & Yap M.G.S. (2007) Recombinant protein therapeutics from CHO cells - 20 years and counting. *Chemical Engineering Progress* 103 ., 40-47.

Jefferis R. (2005) Glycosylation of natural and recombinant antibody molecules. *Adv.Exp.Med.Biol.* 564, 143-148.

Jenuwein T. & Allis C.D. (2001) Translating the histone code. *Science* 293, 1074-1080.

Johnstone R.W. (2002) Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. *Nat.Rev.Drug Discov.* 1, 287-299.

Jones P.A. & Taylor S.M. (1980) Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20, 85-93.

Jones P.L., Veenstra G.J., Wade P.A., Vermaak D., Kass S.U., Landsberger N., Strouboulis J., & Wolffe A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat.Genet.* 19, 187-191.

Kalwy S., Rance J., & Young R. (2006) Toward more efficient protein expression: keep the message simple. *Mol.Biotechnol.* 34, 151-156.

Kaufman R.J. & Sharp P.A. (1982) Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary dna gene. *J.Mol.Biol.* 159, 601-621.

Kellum R. & Schedl P. (1992) A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol.Cell Biol.* 12, 2424-2431.

Kennard,M.L., Goosney,D.L., Monteith,D., Zhang,L., Moffat,M., Fischer,D., and Mott,J. (2009). The generation of stable, high MAb expressing CHO cell lines based on the artificial chromosome expression (ACE) technology. *Biotechnol. Bioeng.* 104, 540-553.

Kim J.M., Kim J.S., Park D.H., Kang H.S., Yoon J., Back K., & Yoon Y. (2004) Improved recombinant gene expression in CHO cells using matrix attachment regions. *J.Biotechnol.* 107, 95-105.

Kim N.S., Kim S.J., & Lee G.M. (1998) Clonal variability within dihydrofolate reductase-mediated gene amplified Chinese hamster ovary cells: stability in the absence of selective pressure. *Biotechnol Bioeng.* 60, 679-688.

Kim S.J., Kim N.S., Ryu C.J., Hong H.J., & Lee G.M. (1998) Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol.Bioeng.* 58, 73-84.

Klose R.J. & Bird A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31, 89-97.

Kohler G. & Milstein C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256, 495-497.

Kosak S.T. & Groudine M. (2004) Form follows function: The genomic organization of cellular differentiation. *Genes and Development* 18, 1371-1384.

Kouzarides T. (2007) Chromatin Modifications and Their Function. *Cell* 128, 693-705.

Kumaki Y., Oda M., & Okano M. (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.* 36, W170-W175.

Kuo M.H. & Allis C.D. (1998) Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* 20, 615-626.

Kuriyama S., Sakamoto T., Kikukawa M., Nakatani T., Toyokawa Y., Tsujinoue H., Ikenaka K., Fukui H., & Tsujii T. (1998) Expression of a retrovirally transduced gene under control of an internal housekeeping gene promoter does not persist due to methylation and is restored partially by 5-azacytidine treatment. *Gene Ther.* 5, 1299-1305.

Kwaks T.H., Barnett P., Hemrika W., Siersma T., Sewalt R.G., Satijn D.P., Brons J.F., van Blokland R., Kwakman P., Kruckeberg A.L., Kelder A., & Otte A.P. (2003) Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nat Biotechnol* 21, 553-558.

Kwaks T.H., Sewalt R.G., van B.R., Siersma T.J., Kasiem M., Kelder A., & Otte A.P. (2005) Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. *J.Biotechnol.* 115, 35-46.

Landry J., Sutton A., Tafrov S.T., Heller R.C., Stebbins J., Pillus L., & Sternglanz R. (2000) The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc.Natl.Acad.Sci.U.S.A* 97, 5807-5811.

Lawrence S. (2005) Biotech drug market steadily expands. *Nat.Biotechnol.* 23, 1466.

Le Hir H., Nott A., & Moore M.J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem.Sci.* 28, 215-220.

Lehnertz B., Ueda Y., Derijck A.A., Braunschweig U., Perez-Burgos L., Kubicek S., Chen T., Li E., Jenuwein T., & Peters A.H. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr.Biol.* 13, 1192-1200.

Lewis J.D., Meehan R.R., Henzel W.J., Maurer-Fogy I., Jeppesen P., Klein F., & Bird A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69, 905-914.

Li E., Bestor T.H., & Jaenisch R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915-926.

Lindahl Allen M. & Antoniou M. (2007) Correlation of DNA methylation with histone modifications across the HNRPA2B1-CBX3 ubiquitously-acting chromatin open element (UCOE). *Epigenetics.* 2, 227-236.

Litt M.D., Simpson M., Recillas-Targa F., Prioleau M.N., & Felsenfeld G. (2001) Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J.* 20, 2224-2235.

Litt M.D., Simpson M., Gaszner M., Allis C.D., & Felsenfeld G. (2001) Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* 293, 2453-2455.

Livak K.J. & Schmittgen T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402-408.

Lucas B.K., Giere L.M., DeMarco R.A., Shen A., Chisholm V., & Crowley C.W. (1996) High-level production of recombinant proteins in CHO cells using a dicistronic DHFR intron expression vector. *Nucleic Acids Res.* 24, 1774-1779.

Luger K., Mader A.W., Richmond R.K., Sargent D.F., & Richmond T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389, 251-260.

Marshall,E. (2001). Rat genome spurs an unusual partnership. Science *291*, 1872.

McBurney M.W., Mai T., Yang X., & Jardine K. (2002) Evidence for repeat-induced gene silencing in cultured Mammalian cells: inactivation of tandem repeats of transfected genes. *Exp.Cell Res.* 274, 1-8.

Meehan R.R., Lewis J.D., McKay S., Kleiner E.L., & Bird A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58, 499-507.

Mermod N. (2007) High-level transgene integration and expression in mammalian cells using MAR elements. Proceedings of the Protein Expression In Animal Cells conference, September 2007.

Meyer P. (2001) Chromatin remodelling. *Curr.Opin.Plant Biol.* 4, 457-462.

Mielke C., Kohwi Y., Kohwi-Shigematsu T., & Bode J. (1990) Hierarchical binding of DNA fragments derived from scaffold-attached regions: correlation of properties in vitro and function in vivo. *Biochemistry* 29, 7475-7485.

Mimura Y., Lund J., Church S., Dong S., Li J., Goodall M., & Jefferis R. (2001) Butyrate increases production of human chimeric IgG in CHO-K1 cells whilst maintaining function and glycoform profile. *J.Immunol.Methods* 247, 205-216.

Mizuguchi G., Tsukiyama T., Wisniewski J., & Wu C. (1997) Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol.Cell* 1, 141-150.

Moehle E.A., Rock J.M., Lee Y.L., Jouvenot Y., DeKelver R.C., Gregory P.D., Urnov F.D., & Holmes M.C. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc.Natl.Acad.Sci.U.S.A* 104, 3055-3060.

Muller H.J. & Altenburg E. (1930) The Frequency of Translocations Produced by X-Rays in Drosophila. *Genetics* 15, 283-311.

Mutskov V.J., Farrell C.M., Wade P.A., Wolffe A.P., & Felsenfeld G. (2002) The barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes and Development* 16, 1540-1554.

Nan X., Meehan R.R., & Bird A. (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* 21, 4886-4892.

Nan X., Ng H.H., Johnson C.A., Laherty C.D., Turner B.M., Eisenman R.N., & Bird A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386-389.

Neuwald A.F. & Landsman D. (1997) GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem.Sci.* 22, 154-155.

Okano M., Xie S., & Li E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat.Genet.* 19, 219-220.

Olins A.L. & Olins D.E. (1974) Spheroid chromatin units (v bodies). *Science* 183, 330-332.

Ooi S.K., Qiu C., Bernstein E., Li K., Jia D., Yang Z., Erdjument-Bromage H., Tempst P., Lin S.P., Allis C.D., Cheng X., & Bestor T.H. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714-717.

Otte A.P., Kwaks T.H., van Blokland R.J., Sewalt R.G., Verhees J., Klaren V.N., Siersma T.K., Korse H.W., Teunissen N.C., Botschuijver S., van M.C., & Man S.Y. (2007) Various expression-augmenting DNA elements benefit from STAR-Select, a novel high stringency selection system for protein expression. *Biotechnol.Prog.* 23, 801-807.

Palermo D.P., DeGraaf M.E., Marotti K.R., Rehberg E., & Post L.E. (1991) Production of analytical quantities of recombinant proteins in Chinese hamster ovary cells using sodium butyrate to elevate gene expression. *J.Biotechnol.* 19, 35-47.

Perucho M., Hanahan D., & Wigler M. (1980) Genetic and physical linkage of exogenous sequences in transformed cells. *Cell* 22, 309-317.

Pikaart M.J., Recillas-Targa F., & Felsenfeld G. (1998) Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* 12, 2852-2862.

Pradhan S., Bacolla A., Wells R.D., & Roberts R.J. (1999) Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J.Biol.Chem.* 274, 33002-33010.

Prendergast G.C. & Ziff E.B. (1991) Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* 251, 186-189.

Prokhortchouk A., Hendrich B., Jorgensen H., Ruzov A., Wilm M., Georgiev G., Bird A., & Prokhortchouk E. (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes and Development* 15, 1613-1618.

Prokhortchouk A., Hendrich B., Jorgensen H., Ruzov A., Wilm M., Georgiev G., Bird A., & Prokhortchouk E. (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes and Development* 15, 1613-1618.

Quina A.S., Buschbeck M., & Di Croce L. (2006) Chromatin structure and epigenetics. *Biochem Pharmacol* 72, 1563-1569.

Recillas-Targa F., Pikaart M.J., Burgess-Beusse B., Bell A.C., Litt M.D., West A.G., Gaszner M., & Felsenfeld G. (2002) Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc.Natl.Acad.Sci.U.S.A* 99, 6883-6888.

Reik W. & Walter J. (2001) Genomic imprinting: parental influence on the genome. *Nat.Rev.Genet.* 2, 21-32.

Reitman M. & Felsenfeld G. (1990) Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken beta-globin locus. *Mol.Cell Biol.* 10, 2774-2786.

Renard J.M., Spagnoli R., Mazier C., Salles M.F., & Mandine E. Evidence that monoclonal antibody production kinetics is related to the integral of the viable cells curve in batch systems.

Rihn B., Coulais C., Bottin M.C., & Martinet N. (1995) Evaluation of non-radioactive labelling and detection of deoxyribonucleic acids. Part One: Chemiluminescent methods. *J.Biochem.Biophys.Methods* 30, 91-102.

Ringold G., Dieckmann B., & Lee F. (1981) Co-expression and amplification of dihydrofolate reductase cDNA and the Escherichia coli XGPRT gene in Chinese hamster ovary cells. *J.Mol.Appl.Genet.* 1, 165-175.

Robertson K.D. & Wolffe A.P. (2000) DNA methylation in health and disease. *Nat.Rev.Genet.* 1, 11-19.

Robinett C.C., O'Connor A., & Dunaway M. (1997) The repeat organizer, a specialized insulator element within the intergenic spacer of the Xenopus rRNA genes. *Mol.Cell Biol.* 17, 2866-2875.

Running D.J. & Allison D.S. (2004) High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1alpha gene. *Biotechnol.Prog.* 20, 880-889.

Sarraf S.A. & Stancheva I. (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol.Cell* 15, 595-605.

Scahill S.J., Devos R., Van der Heyden J., & Fiers W. (1983) Expression and characterization of the product of a human immune interferon cDNA gene in Chinese hamster ovary cells. *Proc.Natl.Acad.Sci.U.S.A* 80, 4654-4658.

Schlesinger Y., Straussman R., Keshet I., Farkash S., Hecht M., Zimmerman J., Eden E., Yakhini Z., Ben-Shushan E., Reubinoff B.E., Bergman Y., Simon I., & Cedar H. (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat.Genet.* 39, 232-236.

Shogren-Knaak M., Ishii H., Sun J.M., Pazin M.J., Davie J.R., & Peterson C.L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844-847.

Simpson D., Williams S., & Irvine A. (2008), Millipore Corporation. Expression Elements. U.S. Pat. 7,632,661

Singh G.B., Kramer J.A., & Krawetz S.A. (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res.* 25, 1419-1425.

Solanas M. & Escrich E. (1997) An improved protocol to increase sensitivity of Southern blot using dig-labelled DNA probes. *J.Biochem.Biophys.Methods* 35, 153-159.

Sorm F., Piskala A., Cihak A., & Vesely J. (1964) 5-Azacytidine, a new, highly effective cancerostatic. *Experientia* 20, 202-203.

Southern E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J.Mol.Biol.* 98, 503-517.

Southern P.J. & Berg P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J.Mol.Appl.Genet.* 1, 327-341.

Sterner D.E. & Berger S.L. (2000) Acetylation of histones and transcription-related factors. *Microbiol.Mol.Biol.Rev.* 64, 435-459.

Stief A., Winter D.M., Stratling W.H., & Sippel A.E. (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* 341, 343-345.

Su X.Z., Wu Y., Sifri C.D., & Wellems T.E. (1996) Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Res* 24, 1574-1575.

Suetake I., Shinozaki F., Miyagawa J., Takeshima H., & Tajima S. (2004) DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J.Biol.Chem.* 279, 27816-27823.

Sun F.L. & Elgin S.C. (1999) Putting boundaries on silence. *Cell* 99, 459-462.

Takai D. & Jones P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc.Natl.Acad.Sci.U.S.A* 99, 3740-3745.

Tamaru H. & Selker E.U. (2001) A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature* 414, 277-283.

Thiagalingam S., Cheng K.H., Lee H.J., Mineva N., Thiagalingam A., & Ponte J.F. (2003) Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann N Y Acad Sci* 983, 84-100.

Tong J.K., Hassig C.A., Schnitzler G.R., Kingston R.E., & Schreiber S.L. (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* 395, 917-921.

Udvardy A., Maine E., & Schedl P. (1985) The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J.Mol.Biol.* 185, 341-358.

Urlaub G. & Chasin L.A. (1980) Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc.Natl.Acad.Sci.U.S.A* 77, 4216-4220.

Urlaub G., Kas E., Carothers A.M., & Chasin L.A. (1983) Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell* 33, 405-412.

Utley R.T., Ikeda K., Grant P.A., Cote J., Steger D.J., Eberharter A., John S., & Workman J.L. (1998) Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* 394, 498-502.

van Blokland H.J., Kwaks T.H., Sewalt R.G., Verhees J.A., Klaren V.N., Siersma T.K., Korse J.W., Teunissen N.C., Botschuijver S., van M.C., Man S.Y., & Otte A.P. (2007) A novel, high stringency selection system allows screening of few clones for high protein expression. *J.Biotechnol.* 128, 237-245.

Van Holde K.E., Sahasrabuddhe C.G., & Shaw B.R. (1974) A model for particulate structure in chromatin. *Nucleic Acids Res.* 1, 1579-1586.

van A.H., Fritsch O., Hohn B., & Gasser S.M. (2004) Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. *Cell* 119, 777-788.

Vignali M., Hassan A.H., Neely K.E., & Workman J.L. (2000) ATP-dependent chromatin-remodeling complexes. *Mol.Cell Biol.* 20, 1899-1910.

Visiongain (2010) Therapeutic Monoclonal Antibodies: World Market 2010-2025. Visiongain Report

Vostrov A.A. & Quitschke W.W. (1997) The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J.Biol.Chem.* 272, 33353-33359.

Wade P.A., Gegonne A., Jones P.L., Ballestar E., Aubry F., & Wolffe A.P. (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat.Genet.* 23, 62-66.

Wade P.A. (2001) Methyl CpG-binding proteins and transcriptional repression. *Bioessays* 23, 1131-1137.

Wallrath L.L. (1998) Unfolding the mysteries of heterochromatin. *Curr.Opin.Genet.Dev.* 8, 147-153.

Walsh G. (2006) Biopharmaceutical benchmarks 2006. *Nat.Biotechnol.* 24, 769-776.

Wang T.Y., Zhang J.H., Jing C.Q., Yang X.J., & Lin J.T. (2010) Positional effects of the matrix attachment region on transgene expression in stably transfected CHO cells. *Cell Biol.Int.* 34, 141-145.

Watt F. & Molloy P.L. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes and Development* 2, 1136-1143.

West A.G., Huang S., Gaszner M., Litt M.D., & Felsenfeld G. (2004) Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol.Cell* 16, 453-463.

Wiebe ME, Beker F, Lazar R, May L, Caste B, Semense M, Fautz C, Gamick R, Miller C, Masover G, Bergmann D & Lubinieeki AS (1989) A multifaceted approach to assure that recombinant tPA is free of adventitious virus. In: *Advances in animal cell biology and technology for bioprocesses*, pp. 68-71. Butterworths, Guildford.

Williams S., Mustoe T., Mulcahy T., Griffiths M., Simpson D., Antoniou M., Irvine A., Mountain A., & Crombie R. (2005) CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol* 5, 17.

Wolffe A.P. & Hayes J.J. (1999) Chromatin disruption and modification. *Nucleic Acids Res.* 27, 711-720.

Wurm F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat.Biotechnol.* 22, 1393-1398.

Wysocka J., Allis C.D., & Coonrod S. (2006) Histone arginine methylation and its dynamic regulation. *Front Biosci.* 11, 344-355.

Yang Y., Mariati, Chusainow J., & Yap M.G. (2010) DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J.Biotechnol.* 147, 180-185.

Ye J., Alvin K., Latif H., Hsu A., Parikh V., Whitmer T., Tellers M., de la Cruz Edmonds MC, Ly J., Salmon P., & Markusen J.F. (2010) Rapid protein production using CHO stable transfection pools. *Biotechnol.Prog.* 26, 1431-1437.

Yoder J.A., Walsh C.P., & Bestor T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335-340.

Yokochi T. & Robertson K.D. (2002) Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. *J.Biol.Chem.* 277, 11735-11745.

Yu,J., Bock,J.H., Slightom,J.L., and Villeponteau,B. (1994). A 5' beta-globin matrix-attachment region and the polyoma enhancer together confer position-independent transcription. *Gene* 139, 139-145.

Zahn-Zabal M., Kobr M., Girod P.A., Imhof M., Chatellard P., de Jesus M., Wurm F., & Mermod N. (2001) Development of stable cell lines for production or regulated expression using matrix attachment regions. *J Biotechnol* 87, 29-42.

Zhang,F., Frost,A.R., Blundell,M.P., Bales,O., Antoniou,M.N., and Thrasher,A.J. (2010). A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. *Mol. Ther.* 18, 1640-1649.

Zhong X.P. & Krangel M.S. (1997) An enhancer-blocking element between alpha and delta gene segments within the human T cell receptor alpha/delta locus. *Proc.Natl.Acad.Sci.U.S.A* 94, 5219-5224.

Zhou J., Chau C.M., Deng Z., Shiekhattar R., Spindler M.P., Schepers A., & Lieberman P.M. (2005) Cell cycle regulation of chromatin at an origin of DNA replication. *EMBO J.* 24, 1406-1417.

APPENDICES

Appendix 1

Chromatin modifying element sequences

1.5kb UCOE

GCGGCCGCACGCGTGGCCCTCCGCGCCTACAGCTCAAGCCACATCCGAAGGGGGAGG

GAGCCGGGAGCTGCGCGCGGGGCCGCCGGGGGGAGGGGTGGCACCGCCCACGCCGGG

CGGCCACGAAGGGCGGGGCAGCGGGCGCGCGCGCGGCGGGGGGAGGGGCCGGCGCCG

CGCCCGCTGGGAATTGGGGCCCTAGGGGGAGGGCGGAGGCGCCGACGACCGCGGCAC

TTACCGTTCGCGGCGTGGCGCCCGGTGGTCCCCAAGGGGAGGGAAGGGGGAGGCGGG

GCGAGGACAGTGACCGGAGTCTCCTCAGCGGTGGCTTTTCTGCTTGGCAGCCTCAGC

GGCTGGCGCCAAAACCGGACTCCGCCCACTTCCTCGCCCGCCGGTGCGAGGGTGTGG

AATCCTCCAGACGCTGGGGGAGGGGGAGTTGGGAGCTTAAAAACTAGTACCCCTTTG

GGACCACTTTCAGCAGCGAACTCTCCTGTACACCAGGGGTCAGTTCCACAGACGCGG

GCCAGGGGTGGGTCATTGCGGCGTGAACAATAATTTGACTAGAAGTTGATTCGGGTG

TTTCCGGAAGGGGCCGAGTCAATCCGCCGAGTTGGGGCACGGAAAACAAAAAGGGAA

GGCTACTAAGATTTTTCTGGCGGGGGTTATCATTGGCGTAACTGCAGGGACCACCTC

CCGGGTTGAGGGGGCTGGATCTCCAGGCTGCGGATTAAGCCCCTCCCGTCGGCGTTA

ATTTCAAACTGCGCGACGTTTCTCACCTGCCTTCGCCAAGGCAGGGGCCGGGACCCT

ATTCCAAGAGGTAGTAACTAGCAGGACTCTAGCCTTCCGCAATTCATTGAGCGCATT

TACGGAAGTAACGTCGGGTACTGTCTCTGGCCGCAAGGGTGGGAGGAGTACGCATTT

GGCGTAAGGTGGGGCGTAGAGCCTTCCCGCCATTGGCGGCGGATAGGGCGTTTACGC

GACGGCCTGACGTAGCGGAAGACGCCTTAGTGGGGGGGAAGGTTCTAGAAAAGCGGC

GGCAGCGGCTCTAGCGGCAGTAGCAGCAGCGCCGGGTCCCGTGCGGAGGTGCTCCTC

GCAGAGTTGTTTCTCCAGCAGCGGCAGTTCTCACTACAGCGCCAGGACGAGTCCGGT

TCGTGTTCGTCCGCGGAGATCTCTCTCATCTCGCTCGGCTGCGGGAAATCGGGCTGA

AGCGACTGAGTCCGCGATGGAGGTAACGGGTTTGAAATCAATGAGTTATTGAAAAGG

GCATGGCGAGGCCGTTGGCGCCTCAGTGGAAGTCGGCCAGCCGCCTCCGTGGGAGAG

AGGCAGGAAATCGGACCAATTCAGTAGCAGTGGGGCTTAAGGTTTATGAACGGGGTC

TTGAGCGGAGGCCTGAGCGTACAAACAGCTTCCCCACCCTCAGCCTCCCGGCGCCAT

TTCCCTTCACTGGGGGTGGGGGATGGGGAGCTTTCACATGGCGGACGCTGCCCCGCT

GGGGTGAAAGTGGGGCGCGGAGGCGGGACTTCTTATTCCCTTTCTAAAGCACGCTGC

TTCGGGGGCCACGGCGTCTCCTCGGACGGCCGGGCGCGCC

MAR X S29

GCGGCCGCACGCGTCTCGAGGATCCCTTTATAAAACCACAATATAATGGAGTGCTAT
AATTTCAAACAGTGTTTGGTCTGCTGGCAGAGTGGTCATTCTAACAGCAGTCACAGT
AGAGTAGAAATAAGACTGCAGTATATCTAAGGCAAAAAGCTGAGGTTTCAGGAGCTT
GAAGGTAAAGAGGAAGAAAGAAATGGGAATGGGAATTGGAAAGACAAATATCGTTAA
GAGAAAATTGCTTTTAGGAGAGGGGAAAGAATCTATGTGTACTTAAGACTATGGAAT
CAATCCCATTTAAGCTGGGAAACTAGTTTCATATATAACTAATAAATTTTATTTACA
GAATATCTATTTACCTGATCTAGGCTTCAAGCCAAAGGGACTGTGTGAAAAACCATC
AGTTCTGTCATATTCCTAAAAAAAAATTAAAAAGTTAAAAATAAATAAATAATAAAA
CTTCTTTTCTTTCAAAATAATCAAGGTGCTTATTCACATCCATTCCAATTTGGGGAA
ATACTTATTTTCCTATGATTAGTGAAGAGAAAAGTAACTTGCATTTCAATTCAAGTT
GATACATGTCACTTTTAAGAGGTCAACTAATATTTGCTAGTTGAGCTAACCATATAG
GCTTTAAATACTTTCATAGTAGAAAGAAAATGAAAATCATTAGTGAACTGTATAAAA
TAGATCATACTTTTTGAAAGAATCAGACTGAAGTTTCCGAAAAAAAGAAGTAAGCTT
CAATGAAAAGGTAAGTGAATTTAGCATTTACTCAGCATCTACTATGGACTTAACACC
TAACAGTAGATAATCTGAAGGCAAACATATTTGTATAGGGACTGCAGAATGATAGAT
GATAAATATCATCTCTTCTATTTGAATGAATATTTTTTCAAATCTTTCACACACAGT
GGTTTGCTATGGAAAGATTTGTAGTACATTAAACAAATCTGAAGATGGAGTTAGAAA
GCTTAGGCTATGTTTTGAGCACAACATATAATTTCTCTGTGATTGTTTCTTCATCTT
TCAAATGAGGTTACTGTGAAGATTAAATGAGATAACTAAATGATGATAAAATAATGT
AATCTTAGCAGCACCTTATTTAATCTGTGCAACAACTCTGTGAAGTGAGTAGGGCTC
AGCTTCAGTCACTTCTCTGCCATTTATTAACTAAGATAGTTTGGAAAGTTACCCATC
TCTTCAGCTGTAAAATGATGAGGATCATACCTATTTTATGGGGCTGCTTTTAGGTAC
AAATATACAGGCAAGCACTTTGTTAATACTAAAGCATTACACCAATTAGTTTTACTC
TTTTCCATTCACACATGAAATTAATGTAATCAGAATTCTGTAGATTACCTAAATCTT
CTGTTAACACGTGATATGCAGTTCAGGTTAAATGTCAGTTGAGTTACCAAAGCACAT
ACATACTCACCACCCTATCCAAATCTACAAGCCTCCCAGTTTGTCTTCACTATTTTG
GTTAAATTAATATGAATTCCTAGATGAAAATTTCACTGATCCAAATGAAATAAAAAA
TATATTACAAAACTCACACCTGTAATCTCAACATTTTGGGAGGCCAAGGCAGGTAGA
TCACTTGAGGCCAGGAGTTCAAGACCAGCCTGATCAACATGGTGAAACCCTGTCTCT
ACTAAAAATACAAAAATTAGCCAGGTGTGGTGGCATGTGCCTGTAGTCCTACCTACT
CGGGAGGCTGAGGCACAAGAATCGCTTGAATGTGGGAGGTGGAGGTTGCAGTGACCT
GAGATCGTGCCACTGCACTCCAGCCTAGGCAACAGAGTGAGATCATGTGTCATATAT

ATATATATATATATATATATATATATATATATATACACACACACACACATATATATA

TACACATATATATACGTATATATATATATATGTATATATATATACATATATATATACATATAT

ATATATACGTATATATATATACGTATATATATATATCAATGTAAATTATTTGGGAAATTTG

GTATGAATAGTCTTCCCTGTGAACACAGATCATAAAATCATATATCAAGCAGACAAA

TAAGTAGTAGTCACTTATATGCTTATACTTGTAACTTAAAGTAAAAGAATTACAAAA

GCATATGACAAAGACTAATTTTAAGATATCCTAATTTAAATTGTTTTCTAAAAGTGT

GTATACCATTTTACCTATCATATGAATAATTTAGAAACATGTTTATAAAATTAATGT

CCAAATCCATTCAAAAGTTTTGTAATGCAGATCACCCACAACAACAAAGAATCCTAG

CCTATTAAAAAAGCAACACCACCTACATATAATGAAATATTAGCAGCATCTATGTAA

CCAAAGTTACACAGTGAATTTGGGCCATCCAACACTTTGAGCAAAGTGTTGAATTCA

TCAAATGAATGTGTAATCATTTACTTACTAATGCCAATACACTTTAAGGTAATCTTA

AGTAGAAGAGATAGAGTTTAGAATTTTTTAAATTTATCTCTTGTTGTAAAGCAATAG

ACTTGAATAAATAAATTAGAAGAATCAGTCATTCAAGCCACCAGAGTATTTGATCGA

GATTTCACAAACTCTAACTTTCTGATACCCATTCTCCCAAAAACGTGTAACCTCCTG

TCGATAGGAACAACCCACTGCAGGGATGTTTCTCGTGGAAAAAGGAAATTTCTTTTG

CATTGGTTTCAGACCTAACTGGTTACAAGAAAAACCAAAGGCCATTGCACAATGCTG

AAGTACTTTTTTCAAATTTAAAATTTGAAAGTTGTTCTTAAAATCTATCATTTATTT

TAAAATACGGATGAATGAGAAAGCATAGATTTGATAAAGTGAATTCTTTTCTGCAAT

CTACAGACACTTCCAAAAATCACTACAGACACTACAGACACTACAGAAAATCATAAA

TAAACAAGTGCTAGTATCAATATTTTTACCAAAAAATGGCATTCTTAGAATTTTTTA

TAGGCTAGAAGGTTTGTACAAACTAATCTGCCACGGATTTTAAAATATGAGTGAATA

AATTATATTGCAAAAAAAATCAGGTTACAGAGAACTGGCAAGGAAGACTCTTATGTA

AAACACAGAAAACATACAAAACGTATTTTTAAGACAAATAAAAACAGAACTTGTACC

TCAGATGATACTGGAGATTGTGTTGACATATTAGCATTATCACTGTCTTGCTAAAAC

ATAAAAATAAAAAGATGGAAGATGAAATTACAATACAAATGATGATTTAAACATATA

AAAGGAAAATAAAAATTGTTCTGACCAACTACTAAAGGAAGACCTACTAAAGATATG

CCATCCAGCACATTGCCACTCTACATGTGGTCTGTAAACCAGCAGCATAGGGATCCA

TCGGCCGGGCGCGCCGTCGAC

STAR 40

GATCAAGAAAGCACTCCGGGCTCCAGAAGGAGCCTTCCAGGCCAGCTTTGAGCATAA

GCTCTGATGAGCAGTGAGTGTCTTGAGTAGTGTTCAGGGCAGCATGTTACCATTCAT

GCTTGACTTCTAGCCAGTGTGACGAGAGGCTGGAGTCAGGTCTCTAGAGAGTTGAGC

AGCTCCAGCCTTAGATCTCCCAGTCTTATGCGGTGTGCCCATTCGCTTTGTGTCTGC
AGTCCCCTGGCCACACCCAGTAACAGTTCTGGGATCTATGGGAGTAGCTTCCTTAGT
GAGCTTTCCCTTCAAATACTTTGCAACCAGGTAGAGAAGTTTGGAGTGAAGGTTTTG
TTCTTCGTTTCTTCACAATATGGATATGCATCTTCTTTTGAAAATGTTAAAGTAAAT
TACCTCTCTTTTCAGATACTGTCTTCATGCGAACTTGGTATCCTGTTTCCATCCCAG
CCTTCTATAACCCAGTAACATCTTTTTTGAAACCAGTGGGTGAGAAAGACACCTGGT
CAGGAACGCGGACCACAGGACAACTCAGGCTCACCCACGGCATCAGACTAAAGGCAA
ACAAGGACTCTGTATAAAGTACCGGTGGCATGTGTATTAGTGGAGATGCAGCCTGTG
CTCTGCAGACAGGGAGTCACACAGACACTTTTCTATAATTTCTTAAGTGCTTTGAAT
GTTCAAGTAGAAAGTCTAACATTAAATTTGATTGAACAATTGTATATTCATGGAATA
TTTTGGAACGGAATACCAAAAAATGGCAATAGTGGTTCTTTCTGGATGGAAGACAAA
CTTTTCTTCTTTAAAATAAATTTTATTTTATATATTTGAGGTTGACCACATGACCTT
AAGGATACATATAGACAGTAAACTGGTTACTACAGTGAAGCAAATTAACATATCTAC
CATCGTACATAGTTACATTTTTTTGTGTGACAGGAACAGCTAAAATCTACGTATTTA
ACAAAACTCCTAAAGACAATACATTTTTATTAACTATAGCCCTCATGATGTACATTA
GATC

STAR 7

GAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTAAAACCTCGTCTCTACTAAA
AAATACGAAAAATTAGCTGGTTGTGGTGGTGCGTGCTTGTAATCCCAGCTACTCGGG
AGGCTGAGGCAGGAGAATCACTTGAATCTGGGAGGCAGAGGTTGCAGTGAGCTGAGA
TAGTGCCATTGCACTCCAGCCTGGGCAACAGACGGAGACTCTGTCTCCAAAAAAAAA
AAAAAAAATCTTAGAGGACAAGAATGGCTCTCTCAAACTTTTGAAGAAAGAATAAAT
AAATTATGCAGTTCTAGAAGAAGTAATGGGGATATAGGTGCAGCTCATGATGAGGAA
GACTTAGCTTAACTTTCATAATGCATCTGTCTGGCCTAAGACGTGGTGAGCTTTTTA
TGTCTGAAAACATTCCAATATAGAATGATAATAATAATCACTTCTGACCCCCCTTTT
TTTTCCTCTCCCTAGACTGTGAAGCAGAAACCCCATATTTTTCTTAGGGAAGTGGCT
ACGCACTTTGTATTTATATTAACAACTACCTTATCAGGAAATTCATATTGTTGCCCT
TTTATGGATGGGGAAACTGGACAAGTGACAGAGCAAAATCCAAACACAGCTGGGGAT
TTCCCTCTTTTAGATGATGATTTTAAAAGAATGCTGCCAGAGAGATTCTTGCAGTGT
TGGAGGACATATATGACCTTTAAGATATTTTCCAGCTCAGAGATGCTATGAATGTAT
CCTGAGTGCATGGATGGACCTCAGTTTTGCAGATTCTGTAGCTTATACAATTTGGTG
GTTTTCTTTAGAAGAAAATAACACATTTATAAATATTAAAATAGGCCCAAGACCTTA
CAAGGGCATTCATACAAATGAGAGGCTCTGAAGTTTGAGTTTGTTCACTTTCTAGTT

AATTATCTCCTGCCTGTTTGTCATAAATGCGTTTAGTAGGGAGCTGCTAATGACAGG
TTCCTCCAACAGAGTGTGGAAGAAGGAGATGACAGCTGGCTTCCCCTCTGGGACAGC
CTCAGAGCTAGTGGGGAAACTATGTTAGCAGAGTGATGCAGTGACCAAGAAAATAGC
ACTAGGAGAAAGCTGGTCCATGAGCAGCTGGTGAGAAAAGGGGTGGTAATCATGTAT
GCCCTTTCCTGTTTTATTTTTTATTGGGTTTCCTTTTGCCTCTCAATTCCTTCTGAC
AATACAAAATGTTGGTTGGAACATGGAGCACCTGGAAGTCTGGTTCATTTTCTCTCA
GTCTCTTGATGTTCTCTCGGGTTCACTGCCTATTGTTCTCAGTTCTACACTTGAGCA
ATCTCCTCAATAGCTAAAGCTTCCACAATGCAGATTTTGTGATGACAAATTCAGCAT
CACCCAGCAGAACTTAGGTTTTTTTCTGTCCTCCGTTTCCTGACCTTTTTCTTCTGA
GTGCTTTATGTCACCTCGTGAACCATCCTTTCCTTAGTCATCTACCTAGCAGTCCTG
ATTCTTTTGACTTGTCTCCCTACACCACAATAAATCACTAATTACTATGGATTCAAT
CCCTAAAATTTGCACAAACTTGCAAATAGATTACGGGTTGAAACTTAGAGATTTCAA
ACTTGAGAAAAAAGTTTAAATCAAGAAAAATGACCTTTACCTTGAGAGTAGAGGCAA
TGTCATTTCCAGGAATAATTATAATAATATTGTGTTTAATATTTGTATGTAACATTT
GAATACCTTCAATGTTCTTATTTGTGTTATTTTAATCTCTTGATGTTACTAACTCAT
TTGGTAGGGAAGAAAACATGCTAAAATAGGCATGAGTGTCTTATTAAATGTGACAAG
TGAATAGATGGCAGAAGGTGGATTCATATTCAGTTTTCCATCACCCTGGAAATCATG
CGGAGATGATTTCTGCTTGCAAATAAAACTAACCCAATGAGGGGAACAGCTGTTCTT
AGGTGAAAACAAAACAAACACGCCAAAAACCTTTATTCTCTTTATTATGAATCAAAT
TTTTCCTCTCAGATAATTGTTTTATTTATTTATTTTTATTATTATTGTTATTATGTC
CAGTCTCACTCTGTCGCCTAAGCTGG

HS4 tandem

GGGGAGCTCACGGGGACAGCCCCCCCCCAAAGCCCCCAGGGATGTAATTACGTCCCT
CCCCCGCTAGGGGGCAGCAGCGACCGCCCGGGGCTCCGCTCCGGTCCGGCGCTCCCC
CCGCATCCCGAGCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGAAGGTGGCACGG
GATCGCTTTCCTCTGAACGCTTCTCGCTGCTCTTTGAGCCTGCAGACACCTGGGGGA
TACGGGGAAAAGGGGAGCTCACGGGGACAGCCCCCCCCCAAAGCCCCCAGGGATGTA
ATTACGTCCCTCCCCCCGCTAGGGGGCAGCAGCGACCGCCCGGGGCTCCGCTCCGGTC
CGGCGCTCCCCCCCGCATCCCGAGCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGA
AGGTGGCACGGGATCGCTTTCCTCTGAACGCTTCTCGCTGCTCTTTGAGCCTGCAGA
CACCTGGGGGGATACGGGGAAAA

Appendix 2.

All oligonucleotides are written 5'-3'

1. Oligonucleotides Used For Cloning

1.1 MAR X_S29:

1.1.1	MAR X_S29 Forward (MAR F)	CGGCCGCACGCGTCTCGAGGATCCCTTTATAAAACCAC *NotI* *MluI* *XhoI*
1.1.2	MAR X_S29 Reverse (MAR R)	GTCGACGGCGCGCCCGGCCGATGGATCCCTATGCTGCTGGTTTACAG *SalI* *AscI* *EagI*

1.2 1.5kb UCOE:

1.2.1	1.5kb UCOE Forward	CTCGAGGGCCCTCCGCGCCTACAGC *XhoI*
1.2.2	1.5kb UCOE Reverse	GTCGACCCGAGGAGACGCCGTGGC *SalI*

1.3 Neomycin cassette:

1.3.1	NEO Forward	GTCGACAAATGAGCTGATTTAAC *SalI*
1.3.2	NEO Reverse	CTCGAGGGTATACAGACATGATAAG *XhoI*

1.4 Glutamine synthetase cassette for STAR 7:

1.4.1	STAR 7 GS Forward	CGGCCGCGTGGGCCTTTCTGCGTTTATAAG *NotI*
1.4.2	STAR 7 GS Reverse	CGGCCGCATTTCGGACAAACACACAT *NotI*

2. Sequencing Oligonucleotides used in cloning

2.1 MAR X_S29 Cloning

2.1.1	1F	GGTGCTTATTCACATCCATTC
2.1.2	2F	TGGTTTGCTATGGAAAG
2.1.3	3F	GATATGCAGTTCAGGTTAAATG
2.1.4	4F	GTGACCTGAGATCGTGCCAC
2.1.5	5F	CTATCATATGAATAATTTAG
2.1.6	6F	CGATAGGAACAACCCACTG
2.1.7	1R	CACGTTTTTGGGAGAATGG
2.1.8	2R	CTAAATTATTCATATGATAGG
2.1.9	3R	GGAGTGCAGTGGCACGATCTC
2.1.10	4R	GACATTTAACCTGAACTGC
2.1.11	5R	CGGAAACTTCAGTCTGATTC
2.1.12	6R	GAATGGATGTGAATAAGCAC

2.2 1.5kb UCOE

2.2.1	1.5kb UCOE F1	CTACAGCTCAAGCCACATCCG
2.2.2	1.5kb UCOE F2	CGAACTCTCCTGTACACCAG
2.2.3	1.5kb UCOE F3	CTCGCAGAGTTGTTTCTCCAG
2.2.4	1.5kb UCOE R1	CACCTTTCAGGGGTCCGAGG
2.2.5	1.5kb UCOE R2	CTCATTGATTTCAAACCCG
2.2.6	1.5kb UCOE R3	CGCTCAATGAATTGCGGAAG

2.3 Neomycin cassette

2.3.1	NEO F1	CAACGGCCTCAACCTACTACTG
2.3.2	NEO R2	GCACGTACTCGGATGGAAGC
2.3.3	NEO R1	CCTGTCTCTTGATCAGAGCTTG
2.3.4	NEO R2	GCTTCCATCCGAGTACGTGC

2.4 Glutamine synthetase cassette cloning for STAR 7

2.4.1	STAR7 GS For	GCACTGGGGACAGCCTATTTTGCTAG
2.4.2	GS NotI F1	CACAAAAAACCAACACACAGATGTA
2.4.3	GS NotI F2	GAGAAAGTCCAAGCCATGTATATC
2.4.4	GS NotI F3	CAAATGCTGAGGTCATGCCT
2.4.5	GS NotI F4	GCCTTCTCAATGAGACTGGC
2.4.6	GS NotI F5	AAGCTGCACTGCTATACAAGAAA
2.4.7	STAR7 GS rev	AGGCCCACCAATATCGCCATC
2.4.8	GS NotI R1	CGGAAATGTTGAATACTCATACTCT
2.4.9	GS NotI R2	GTAACCATTATAAGCTGCAA
2.4.10	GS NotI R3	TGGGGTCTTCTACCTTTCTCTTC
2.4.11	GS NotI R4	GCCGCGGTCTTCAAAGTAAC
2.4.12	GS NotI R5	CTGTAATCTTGACCCCAGCAT

3. TaqMan probes for qPCR

3.1 Hamster GAPDH Forward	CTGCCACCCAGAAGACTGT
3.2 Hamster GAPDH Reverse	GTGGATGCAGGGATGATGTTCT
3.3 Hamster GAPDH Reporter	ATCACGCCACAGCTTT
3.4 Murine Kappa Forward	GGAAGATTGATGGCAGTGAACGA
3.5 Murine Kappa Reverse	GCTGTCCTGATCAGTCCAACT
3.6 Murine Kappa Reporter	TCAGGACGCCATTTTG
3.7 Murine Heavy Forward	GAGCAGTTCAACAGCACTTTCC
3.8 Murine Heavy Reverse	GCCAGTCCTGGTGCATGAT
3.9 Murine Heavy Reporter	CAGTCAGTGAACTTCC

4. Southern blot hybridization probe generation

4.1 Forward	GTCATGAGATTATCAAAAAGGATC
4.2 Reverse	CGCCTTGATCGTTGGGAACC

5. Bisulfite sequencing

5.1 BSPLF1	TTATAGGTGTGGGTTATTGATT
5.2 BSPLR1	AAAAAATAACCACCCTAACAAT
5.3 BSPLF2	TTTTTATAGGATGGGGTTTTATT
5.4 BSPLR2	CACATCTAACACCTAAAAACCA
5.5 BSPHF1	GTTATATTGTTTTTGGTTTGGG
5.6 BSPHR1	TACTCCAAACCCTTCTCAAATAC
5.7 BSPHF2	TTTTGTATTTTTATAGGATGGGGT
5.8 BSPHR2	ACTCCCTCCAAATTTAACAAAC

Appendix 3. Distribution of expressing clones

Clones isolated from *neo* pooled stables.

1.5kb UCOE

	Plate	No growth	1st quarter	2nd quarter	3rd quarter	4th quarter	Total Clones
Ab535	1	50	2 4.3	0 0.0	2 4.3	42 91.3	46
	2	0	2 2.1	8 8.3	1 1.0	85 88.5	96
3' HC	1	15	4.0 4.9	6.0 7.4	23 28.4	48 59.3	81
	2	25	2.0 2.8	4.0 5.6	9 12.7	56 78.9	71
3' HC 5' LC	1	4	19 20.7	12 13.0	8 8.7	53 57.6	92
	2	36	10 16.7	4 6.7	6 10.0	40 66.7	60
3' HC 5' HC	1	7	14 15.7	21 23.6	15 16.9	39 43.8	89
	2	9	13 14.9	11 12.6	12 13.8	51 58.6	87
3' HC 5' LC 5' HC	1	15	25 30.9	6 7.4	11 13.6	39 48.1	81
	2	8	22 25.0	10 7.4	10 7.4	46 52.3	88
5' LC	1	15	18 22.2	14 17.3	5 6.2	44 54.3	81
	2	21	4 5.3	13 17.3	4 5.3	54 72.0	75
5' HC	1	7	3 3.4	4 4.5	4 4.5	78 87.6	89
	2	13	7.0 8.4	3.0 3.6	6.0 7.2	67 80.7	83
5' LC 5' HC	1	17	19 24.1	4 5.1	8 10.1	48 60.8	79
	2	17	12 15.2	10 12.7	9 11.4	48 60.8	79

MAR X S29

	Plate	No growth	1st quartile	2nd quartile	3rd quartile	4th quartile	Total Clones
Ab535	1	10	5	5	8	68	86
			5.8	5.8	9.3	79.1	
	2	19	1	7	7	62	77
			1.3	9.1	9.1	80.5	
3' HC	1	8	3	3	10	72	88
			3.4	3.4	11.4	81.8	
	2	54	2	3	3	34	42
			4.8	7.1	7.1	81.0	
3' HC 5' LC	1	18	7	8	9	54	78
			9.0	10.3	11.5	69.2	
	2	12	10	7	11	56	84
			11.9	8.3	13.1	66.7	
3' HC 5' HC	1	18	4	6	10	58	78
			5.1	7.7	12.8	74.4	
	2	17	2	2	9	66	79
			2.5	2.5	11.4	83.5	
3' HC 5' LC 5' HC	1	8	11	4	6	67	88
			12.5	4.5	6.8	76.1	
	2	27	6	9	7	47	69
			8.7	13.0	10.1	68.1	
5' LC	1	12	2	3	12	67	84
			2.4	3.6	14.3	79.8	
	2	22	4	9	10	51	74
			5.4	12.2	13.5	68.9	
5' HC	1	16	5	10	5	60	80
			6.3	12.5	6.3	75.0	
	2	24	4	5	4	59	72
			5.6	6.9	5.6	81.9	
5' LC 5' HC	1	10	4	6	9	67	86
			4.7	7.0	10.5	77.9	
	2	29	2	1	9	55	67
			3.0	8.8	5.9	70.6	

STAR 40

	Plate	No growth	1st quartile	2nd quartile	3rd quartile	4th quartile	Total Clones
Ab535	1	15	2	2	9	68	81
			2.5	2.5	11.1	84.0	
	2	9	4	5	8	70	87
			4.6	5.8	9.2	80.5	
3' HC	1	17	5	6	2	66	79
			6.3	7.6	2.5	83.5	
	2	19	5	1	3	68	77
			6.5	1.3	3.9	88.3	
3' HC 5' LC	1	18	5	1	4	68	78
			6.4	1.3	5.1	87.2	
	2	14	4	4	4	70	82
			4.9	4.9	4.9	85.4	
3' HC 5' HC	1	38	3	1	4	50	58
			5.2	1.7	6.9	86.2	
	2	15	3	2	8	68	81
			3.7	2.5	9.9	84.0	
3' HC 5' LC 5' HC	1	29	1	1	3	62	67
			1.5	1.5	4.5	92.5	
	2	0	7	2	3	84	96
			7.3	2.1	3.1	87.5	
5' LC	1	9	7	4	7	69	87
			8.0	4.6	8.0	79.3	
	2	30	2	2	4	58	66
			3.0	3.0	6.1	87.9	
5' HC	1	4	3	9	5	75	92
			3.3	9.8	5.4	81.5	
	2	3	5	4	8	76	93
			5.4	4.3	8.6	81.7	
5' LC 5' HC	1	28	2	7	3	56	68
			2.9	10.3	4.4	82.4	
	2	20	2	4	0	70	76
			2.9	5.3	0.0	92.1	

HS4 tandem

	Plate	No growth	1st quartile	2nd quartile	3rd quartile	4th quartile	Total Clones
Ab535	1	8	3	4	4	77	88
			3.4	4.5	4.5	87.5	
	2	20	2	0	3	71	76
			2.6	0.0	3.9	93.4	
3' HC	1	4	3	2	8	79	92
			3.3	2.2	8.7	85.9	
	2	14	4	6	6	66	82
			4.9	7.3	7.3	80.5	
3' HC 5' LC	1	6	2	1	6	81	90
			2.2	1.1	6.7	90.0	
	2	5	1	3	3	84	91
			1.1	3.3	3.3	92.3	
3' HC 5' HC	1	4	4	3	5	80	92
			4.3	3.3	5.4	87.0	
	2	8	2	1	8	77	88
			2.3	1.1	9.1	87.5	
3' HC 5' LC 5' HC	1	23	3	1	6	63	73
			4.1	1.4	8.2	86.3	
	2	25	2	2	5	62	71
			2.8	2.8	7.0	87.3	
5' LC	1	12	3	7	8	66	84
			3.6	8.3	9.5	78.6	
	2	51	1	1	5	44	45
			2.2	2.2	11.1	97.8	
5' HC	1	12	3	5	4	72	84
			3.6	6.0	4.8	85.7	
	2	36	2	2	5	51	60
			3.3	3.3	8.3	85.0	
5' LC 5' HC	1	6	9	2	6	73	90
			10.0	2.2	6.7	81.1	
	2	20	3	4	7	62	76
			3.9	8.8	5.9	70.6	

Clones isolated by limiting dilution under GS selection

	1st quarter	2nd quarter	3rd quarter	4th quarter	Total Clones
Ab535	15 19.2	5 6.4	2 2.6	56 71.8	78
1.5kb UCOE	97 78.9	2 1.6	2 1.6	22 17.9	123
MAR X_S29	22 41.5	5 9.4	2 3.8	24 45.3	53
STAR 7	17 21.0	3 3.7	6 7.4	55 67.9	81
HS4 Tandem	23 38.3	1 1.7	1 1.7	35 58.3	60

Clones isolated from GS pooled stables

	No growth	1st quarter	2nd quarter	3rd quarter	4th quarter	Total Clones
Ab535	28	45 27.4	10 6.1	4 2.4	105 64.0	164
1.5kb UCOE	14	143 80.3	27 15.2	6 3.4	2 1.1	178
MAR X_S29	5	59 31.6	14 7.5	5 2.7	109 58.3	187
STAR 7	42	11 7.3	4 2.7	4 2.7	131 87.3	150
HS4 Tandem	23	53 31.4	17 10.1	8 4.7	144 85.2	169

Appendix 4: Copy number determination by Q-PCR

Ab535

Number of Cells

Sample Name	Ct	Quantity	Avg quantity (ng)	# cells
535 1	17 79	150 56	225 60	14182 22
	17 14	258 58		
	17 10	267 67		
535 2	18 65	73 81	70 11	10622 44
	18 63	75 16		
	18 88	61 35		
535 3	17 61	174 78	163 73	24807 64
	17 80	150 05		
	17 67	166 37		
535 4	18 04	122 34	133 38	20208 76
	17 93	134 57		
	17 85	143 22		
535 5	17 72	160 56	195 89	29680 30
	17 57	181 41		
	17 20	245 70		
535 6	13 42	227 73	205 85	31188 69
	13 88	178 87		
	13 57	210 94		
535 7	17 46	198 64	224 80	34061 05
	17 63	171 83		
	16 95	303 94		
535 8	18 25	103 15	115 26	17464 23
	18 15	112 41		
	17 97	130 23		
535 9	18 34	95 38	113 66	17221 21
	18 18	109 19		
	17 91	136 41		
535 10	18 70	71 08	72 39	10968 70
	18 61	76 99		
	18 72	69 71		

HC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
535 1	28 98	1126 81	973 96	139 86	0 03	0 99
	29 25	942 71				
	29 40	852 36				
535 2	27 81	2475 68	2354 12	110 51	0 22	0 70
	27 90	2326 95				
	27 94	2259 73				
535 3	25 43	12200 58	12361 14	304 09	0 50	0 47
	25 43	12174 87				
	25 36	12713 96				
535 4	24 35	25115 92	23041 40	1817 64	1 14	1 08
	24 51	22279 95				
	24 56	21728 32				
535 5	25 36	12741 96	19068 88	4327 95	0 51	0 60
	25 40	12402 17				
	24 68	20062 51				
535 6	26 58	12476 24	13272 13	691 87	0 43	0 49
	26 45	11610 18				
	26 44	13729 98				
535 7	23 50	44188 62	48976 38	9571 28	1 64	1 34
	23 55	42743 72				
	23 05	59996 81				
535 8	24 60	21170 44	21645 68	1410 72	1 34	0 17
	24 65	20526 54				
	24 46	23242 05				
535 9	25 94	8643 85	9556 90	1149 82	0 55	0 82
	25 85	9184 85				
	25 80	10842 46				
535 10	26 98	6416 45	6171 81	652 90	0 98	1 61
	26 93	6668 85				
	26 83	5432 33				

LC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
535 1	28 08	2901 42	2646 93	160 86	0 08	0 00
	28 11	2542 28				
	27 96	2807 10				
535 2	27 96	2838 86	2997 19	190 42	0 28	0 01
	27 83	3091 06				
	27 81	3101 63				
535 3	25 94	10770 92	10287 57	649 28	0 41	0 03
	25 98	10097 08				
	26 12	9914 71				
535 4	24 88	21790 82	22666 58	993 79	1 11	0 05
	24 81	22877 58				
	24 81	23743 33				
535 5	24 63	29804 63	28884 97	1038 80	0 91	0 17
	25 03	27549 33				
	26 36	29301 10				
535 6	26 14	8194 61	8791 29	1931 41	0 28	0 09
	25 83	10130 83				
	26 33	7818 78				
535 7	23 17	68086 92	68098 77	7126 70	1 02	0 21
	23 00	62310 00				
	22 89	78184 80				
535 8	24 93	27490 86	27986 64	8987 21	1 96	0 60
	24 14	33731 80				
	25 11	18720 12				
535 9	26 17	9547 31	9934 82	1004 31	0 86	0 00
	26 96	7113 79				
	26 57	13084 36				
535 10	26 98	5470 86	5873 43	494 64	0 54	0 08
	26 73	6179 11				
	26 85	5970 18				

1.5kb UCOE

Number of Cells

Sample Name	Ct	Quantity	Avg quantity (ng)	# cells
U1	17 55	144 52	149 52	22655 10
	17 49	151 82		
	17 49	152 23		
U2	18 71	58 27	60 68	9193 38
	18 69	59 39		
	18 58	64 37		
U3	18 48	70 01	65 52	9927 07
	18 58	64 49		
	18 63	62 06		
U4	18 32	79 13	75 55	11446 86
	18 37	76 12		
	18 45	71 40		
U5	19 42	33 31	32 93	4928 87
	19 33	35 77		
	19 62	28 51		
U6	19 04	84 04	86 13	13050 49
	18 92	90 55		
	19 05	83 80		
U7	19 39	34 28	37 45	5673 63
	19 26	37 95		
	19 19	40 11		
U8	17 57	142 81	165 69	25104 22
	17 43	159 86		
	17 18	194 39		
U9	17 37	166 95	158 20	23970 08
	17 41	161 93		
	17 54	145 74		
U10	17 21	189 37	193 32	29291 35
	16 94	234 19		
	17 45	156 41		

HC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
U1	25 39	15138 23	12715 82	2124 93	0 56	0 09
	25 74	11842 64				
	25 82	11166 57				
U2	24 08	38888 34	42320 28	3177 76	4 60	0 35
	23 85	45160 77				
	23 92	42911 73				
U3	25 39	15158 82	14430 46	692 83	1 45	0 07
	25 47	14352 86				
	25 53	13779 70				
U4	23 78	47580 11	42330 82	4867 27	3 70	0 43
	23 97	41445 14				
	24 10	37967 21				
U5	27 89	2576 13	2603 42	52 62	0 53	0 01
	27 85	2664 06				
	27 90	2569 87				
U6	25 37	17066 45	17212 61	681 30	1 32	0 01
	25 48	16633 95				
	25 45	17037 44				
U7	24 89	21614 51	21178 42	585 12	3 75	0 10
	24 91	21603 08				
	24 97	20511 66				
U8	24 68	25065 34	29939 52	13861 94	1 10	0 55
	25 06	19173 00				
	23 84	45580 21				
U9	24 92	21259 44	28618 31	10500 36	1 10	0 44
	24 75	21943 87				
	24 00	40619 62				
U10	24 37	21250 56	33729 12	11461 08	1 15	0 39
	24 76	21710 12				
	23 82	46226 68				

LC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
U1	26 75	10076 79	11309 18	1628 80	0 49	0 07
	26 70	10468 79				
	26 58	13382 96				
U2	24 27	58279 18	58941 43	3194 01	8 61	0 55
	24 32	56130 44				
	24 17	62414 67				
U3	26 00	17084 32	17510 90	449 97	1 77	0 05
	25 97	17524 19				
	25 93	17984 19				
U4	23 98	71663 91	72765 88	1368 86	6 98	0 11
	23 93	73949 18				
	23 93	72684 57				
U5	28 55	3821 58	3622 08	363 61	0 81	0 09
	28 69	3991 61				
	28 83	3293 19				
U6	25 83	21 923 90	20847 61	987 60	1 98	0 09
	25 71	20660 94				
	25 73	20189 72				
U7	24 87	58160 77	58219 05	1929 69	6 54	0 56
	24 96	56329 69				
	24 79	60186 68				
U8	23 48	101 689 10	83270 33	16661 88	3 69	1 56
	23 25	81911 84				
	23 60	66682 68				
U9	24 98	33438 71	29138 96	17921 66	3 33	0 56
	24 87	36181 93				
	26 22	18075 63				
U10	23 08	57871 58	39668 13	6110 66	3 01	0 16
	23 67	36979 61				
	23 16	34985 17				

STAR 7

Number of Cells

Sample Name	Ct	Quantity	Avg quantity (ng)	# cells
S1	18 79	60 89	61 48	9315 47
	18 74	62 45		
	18 78	61 10		
S2	18 50	71 99	64 52	9775 57
	18 44	74 47		
	19 23	47 09		
S3	17 65	117 89	114 71	17380 08
	17 67	115 90		
	17 76	110 33		
S4	16 91	180 53	179 74	27233 01
	16 90	181 07		
	16 94	177 61		
S5	17 86	104 42	111 98	16966 52
	17 62	119 80		
	17 74	111 72		
S6	17 95	98 82	101 11	15320 28
	17 82	106 76		
	17 97	97 76		
S7	18 48	72 83	69 58	10542 09
	18 62	67 20		
	18 58	68 70		
S8	19 89	32 11	32 98	4996 40
	19 82	33 54		
	19 84	33 08		
S9	18 43	75 02	76 44	11582 38
	18 39	76 62		
	18 37	77 68		
S10	18 91	56 22	62 94	9537 01
	18 60	68 02		
	18 69	64 60		

HC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
S1	19 05	643092 30	661499 80	41420 40	71 01	4 45
	19 08	632473 80				
	18 91	708933 50				
S2	22 92	47448 17	48115 05	5718 11	4 92	0 58
	22 73	54137 37				
	23 08	42759 61				
S3	21 39	131517 40	127182 56	15928 79	7 32	0 92
	21 33	138969 36				
	21 69	109060 93				
S4	30 58	272 06	285 32	12 27	0 01	0 00
	30 50	287 64				
	30 45	296 28				
S5	23 88	24918 15	27691 49	2452 15	1 63	0 97
	23 62	29572 60				
	23 67	28583 71				
S6	21 75	104597 48	114187 91	14868 10	7 45	0 14
	21 72	108651 15				
	21 41	131315 12				
S7	21 09	42189 64	41878 53	1707 73	1 97	0 18
	21 06	43272 15				
	21 18	39971 60				
S8	19 52	470114 18	485914 16	22918 32	97 29	4 59
	19 50	475403 66				
	19 39	512224 40				
S9	22 70	95203 09	73551 98	17163 17	6 35	1 58
	22 23	79728 63				
	21 98	89724 17				
S10	22 11	81993 77	77265 04	8923 00	8 10	0 94
	22 10	82827 35				
	22 41	66974 01				

LC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
S1	19 22	716651 80	721886 10	41687 61	77 09	4 99
	19 16	670641 44				
	19 18	778361 75				
S2	21 40	49281 22	52457 39	8081 47	5 51	0 83
	21 23	50870 22				
	22 05	61214 43				
S3	21 01	121108 79	120125 01	2082 91	6 91	0 13
	21 07	121487 82				
	21 07	117779 11				
S4	30 72	340 11	320 97	37 89	0 01	0 00
	30 75	311 99				
	30 87	309 81				
S5	23 73	16211 06	18615 18	7611 21	2 70	0 84
	23 55	16044 64				
	23 09	22586 95				
S6	21 49	186761 36	169967 23	14311 86	9 70	1 11
	21 55	177931 27				
	22 08	145142 06				
S7	23 43	66631 77	69416 58	6962 86	4 11	0 61
	23 23	80738 91				
	23 55	60979 41				
S8	19 10	864 719 10	613629 00	61991 10	173 91	13 00
	19 00	544318 90				
	19 06	630 849 90				
S9	21 07	11 7911 30	1060931 07	91975 30	9 09	0 93
	21 02	1 000 34 09				
	23 31	87964 76				
S10	22 83	93135 11	98100 71	11056 43	9 84	1 26
	22 33	87951 87				
	22 92	73331 16				

HS4 Tandem

Number of Cells

Sample Name	Ct	Quantity	Avg quantity (ng)	# cells
H1	17 71	153 04	151 24	22915 49
	17 73	151 11		
	17 74	149 57		
H2	18 18	110 07	136 34	20638 17
	17 82	141 80		
	17 67	157 16		
H3	18 52	86 33	82 46	12494 55
	18 59	82 33		
	18 66	78 74		
H4	17 71	152 94	145 71	22076 96
	17 87	136 77		
	17 76	147 41		
H5	18 60	81 79	78 86	11949 19
	18 68	77 52		
	18 68	77 29		
H6	18 78	71 97	73 08	11073 42
	18 71	75 66		
	18 79	71 61		
H7	18 69	76 97	73 36	11115 62
	18 80	71 22		
	18 78	71 90		
H8	18 75	73 59	70 40	10666 28
	18 87	67 94		
	18 83	69 66		
H9	18 23	105 99	110 35	16719 86
	18 19	108 75		
	18 10	116 31		
H10	18 37	95 88	100 74	15263 15
	18 30	101 29		
	18 24	105 04		

HC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
H1	24 24	18089 97	17963 83	874 17	0 78	0 04
	24 18	18768 08				
	24 33	17033 44				
H2	25 19	9426 39	10220 54	1254 57	0 49	0 06
	24 88	11642 87				
	25 17	9592 36				
H3	24 05	20607 80	18281 00	2068 10	1 46	0 17
	24 36	16652 26				
	24 28	17582 94				
H4	21 68	103280 32	118577 27	15649 98	5 39	0 71
	21 49	117893 62				
	21 29	134557 88				
H5	23 34	13147 71	24518 54	7968 02	2 05	0 67
	24 26	17862 61				
	23 91	22345 29				
H6	23 44	31109 45	31371 09	855 81	2 83	0 08
	23 39	32327 17				
	23 46	30676 64				
H7	24 65	15683 08	13795 26	359 10	1 24	0 03
	24 67	11505 54				
	24 90	14197 35				
H8	24 95	11154 51	10118 43	985 96	0 91	0 09
	25 11	9958 15				
	25 22	9242 63				
H9	21 34	110286 96	128964 12	5371 89	7 71	0 32
	21 42	123052 65				
	21 30	133553 33				
H10	25 19	9444 66	10158 02	1435 72	0 66	0 09
	25 23	9180 59				
	24 87	11788 82				

LC

Sample Name	Ct	Quantity	Qty Mean	Qty StdDev	Copy per cell	Copy per cell SD
H1	24 99	12081 96	11849 12	411 83	0 92	0 03
	24 97	12163 88				
	25 07	11580 51				
H2	24 60	17797 91	17932 43	2658 76	0 83	0 13
	24 22	20048 49				
	24 68	14750 88				
H3	24 47	16986 56	17173 99	2110 96	1 30	0 17
	24 61	15477 10				
	24 25	19061 10				
H4	22 36	44827 82	45983 45	22187 76	2 08	1 08
	23 12	42048 66				
	23 83	26013 87				
H5	25 08	11478 60	10410 09	1470 28	0 87	0 12
	25 46	8711 91				
	25 12	11020 15				
H6	23 81	26023 23	25699 66	982 54	3 93	0 39
	23 93	24611 30				
	23 81	26464 91				
H7	24 69	14636 64	15065 18	3545 48	1 41	0 33
	24 56	18347 07				
	24 76	14011 69				
H8	25 03	11766 81	10951 51	716 98	1 03	0 07
	24 90	10419 10				
	25 17	10668 73				
H9	22 18	29819 30	30098 77	8003 88	4 65	0 68
	23 02	24081 47				
	22 90	31615 55				
H10	24 13	21561 10	19161 23	6106 91	1 27	0 41
	23 91	24815 92				
	24 05	15206 70				